databricks

eBook

Accelerate Industry Innovation with Data Sharing

Best practices for secure collaboration across clouds, platforms, and regions

Contents

CHAPTER 1

Executive summary: The impact of open data sharing across industries

The underlying challenge for all industries is also the same thing fueling opportunity — data. Discovering available data is one thing, but the journey from discovery to realizing outcomes can be slow and costly. After years of adding various data products for different teams and use cases, top-performing enterprises are now embracing the unification of their data ecosystem to break down data silos, enable cross-team collaboration, and inform smart decision-making for measurable business outcomes.

> The potential benefit is palpable: Gartner's survey of CDOs found that organizations that promote external data sharing have three times the measurable economic benefit across a variety of performance metrics compared to their peers.
>
> "Data Sharing Is a Business Necessity to Accelerate Digital Business"
> Gartner, 2021

Enterprise leadership and technology decision-makers can utilize this eBook to understand the challenges preventing secure and efficient data sharing for data, analytics, and AI workload optimization. Learn about established and emerging data sharing best practices by industry with real-world examples using the Databricks Data Intelligence Platform.

Discover how various industries, including healthcare and life sciences, financial services, retail and consumer goods, communications, media and entertainment, manufacturing and energy, and the public sector use the Databricks Data Intelligence Platform's native and integrated tools for data sharing while employing effective data governance and privacy-safe security. See how the open Databricks Marketplace can accelerate analytics and AI initiatives with assets like AI models, pre-built notebooks, applications, and dashboards — without proprietary vendor lock-in or legacy technology limitations. With more than 100 data providers and 1,000 listings, Databricks Marketplace is supported by a fast-growing ecosystem of data providers with abundant opportunities for business expansion through data and analytics use cases across industries such as:

- **Drug discovery:** Facilitate collaborative analysis of diverse datasets across teams and organizations to identify patterns, uncover potential targets, and optimize therapeutic approaches for improved healthcare outcomes.
- **Process optimization:** Gain complete visibility across the value chain to balance real-time inventory against multi-faceted supply forecasts, predict and implement maintenance services before outages, and maintain quality control with efficient resource allocation.
- **Open finance:** Make financial data easily accessible and shareable to foster innovation and improved consumer experiences.
- **Targeting audiences:** Understand how to boost customer engagement with next-best actions based on individual behaviors and statistical likelihoods generated from external data.

An open ecosystem to help accelerate innovation

Open Source Clients

Apache Spark, pandas, presto, trino, R, Apache Hive

Commercial Clients

Business Intelligence

tableau, Power BI, Qlik, Looker

Analytics

databricks, Google Big Query, Microsoft Azure, dremio, ORACLE Cloud

Governance

IMMUTA, privacera, Alation, collibra

Data Providers

Nasdaq, S&P Global, ICE, NYSE, precisely, SAFEGRAPH, ATLASSIAN, FOURSQUARE, sequence bio, veraset, rearc, dun & bradstreet, ADP, ACXIOM, CARTO, AccuWeather, LexisNexis, Facteus

CHAPTER 2

Exploring the applications of data sharing across industries

Legacy data sharing is costly and complex

Until recently, data sharing was an obstacle for enterprises due to inherent data silos, privacy concerns, and access restrictions that create barriers for internal and external stakeholders. Many organizations are juggling various data sources across disparate platforms, making it nearly impossible to advance data initiatives with efficiency and reliability.

Legacy data sharing looks like this: companies exchange data through outdated methods like, FTP and SSH. While these methods worked in the past, as technology and data have evolved, legacy sharing becomes time-consuming and expensive.

Data extracts and SFTP sharing often require up to five different systems to maintain (spanning database, ETL, SFTP, job management, monitoring sources). Each data transfer job then requires a project to create, test and implement data sharing between companies. And lastly, with any data replication solution, over time, the data could easily become out of sync.

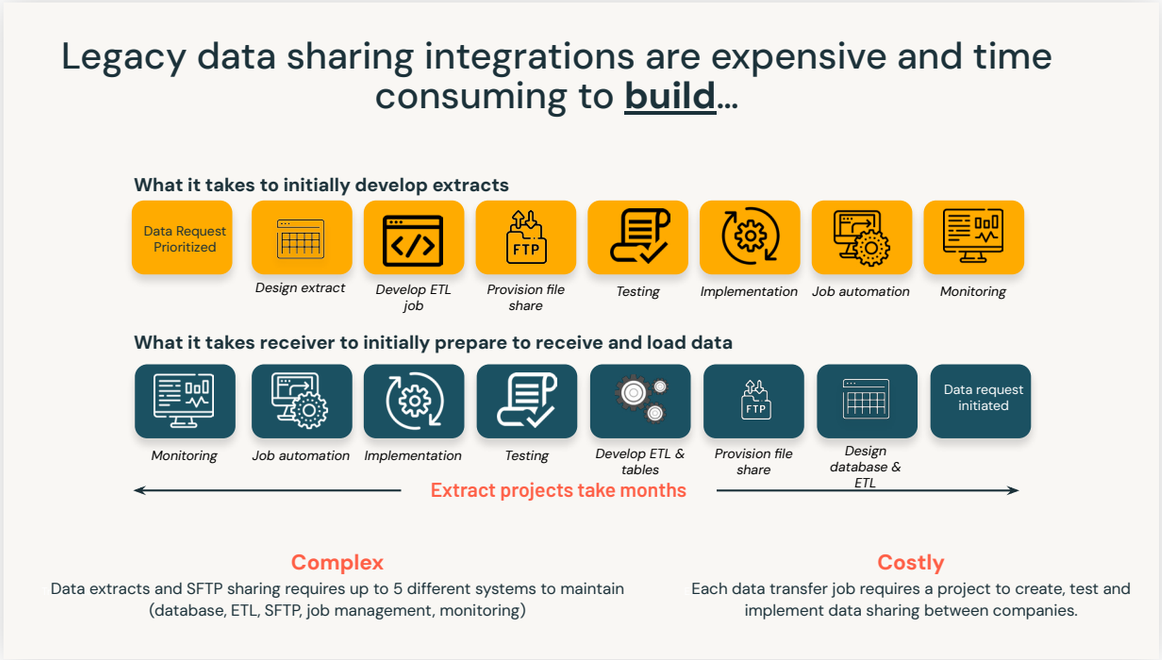
What is data sharing?

Data sharing refers to distributing or providing access to digital information for individuals, organizations, or systems with a legitimate need for that data. Typically, data is shared in the form of data sets, files, or databases and involves granting permissions or sharing protocols that allow authorized users to retrieve, use, or contribute to the shared data.

Data sharing can take various forms, from enabling collaboration between organizations to making research findings widely available, to delivering data to other organizations, or to support proprietary data.

When done effectively, data sharing accelerates internal and external team collaboration, enables analysis, drives innovation, and enhances transparency while adhering to privacy and security considerations to protect sensitive and confidential information.

> In a recent Chief Data Officer survey, **Gartner found that organizations who promote external data sharing have three times the measurable economic benefit** across a variety of performance metrics compared to their peers.
>
> GARTNER



That leaves organizations with unnecessary cost and complexity, impacting how fast they can bring together teams and new products to market.

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

The future of industry is built on collaboration

Data has emerged as the lifeblood of innovation driving progress across industries. The increase in volume and higher quality of data available has enabled better decision-making and resulted in groundbreaking innovations. The massive opportunity across every industry is transforming how data is harnessed — fueling innovation while creating a safe and controlled environment for stakeholders to collaborate and engage data to optimize applications, use cases, and outputs across industries.

Open for collaboration: data sharing made simple on the Databricks Data Intelligence Platform

Databricks empowers enterprises to share live data across data platforms, clouds or regions without friction or limitations, and efficiently govern, track, and audit access to shared data sets. The engine powering this sharing: Delta Sharing.

Delta Sharing is the first open source approach to data sharing across data, analytics, and AI, simplifying data sharing with other organizations regardless of your computing platform. So, no matter how many data partners your organization plans to share with, the pace of your innovation isn't slowed down by vendor lock-in.

Databricks recently expanded the Delta Sharing ecosystem with new partners, including Cloudflare, Dell, Oracle and Twilio to seamlessly share data between their platforms, Databricks, Apache Spark™, pandas, Power BI, Excel and any other system that supports the open protocol.

databricks

Data Sharing with Databricks at a Glance

- **MARKETPLACE:** An open marketplace that allows you to easily discover, evaluate and gain access to data products including data sets, machine learning (ML) models, dashboards and notebooks from anywhere, without the need to be on the Databricks Platform.
- **CLEAN ROOMS:** A secure and privacy-safe environment for enterprises to collaborate and securely share data from your data lakes without data replication.
- **UNITY CATALOG:** The industry's first unified governance solution, enabling organizations to seamlessly govern their structured and unstructured data, machine learning models, notebooks, dashboards and files on any cloud or platform.

Instant data delivery, zero ETL costs with Databricks Marketplace

Organizations can get started with instant delivery of data with partners in the ecosystem today with Databricks Marketplace. Databricks Marketplace is an open marketplace for data, analytics, and AI, powered by the open source Delta Sharing standard. Bringing together a vast ecosystem of data consumers and data providers, the Marketplace enables collaboration with a wide array of data and AI products, including data sets, AI models, pre-built notebooks, applications, and dashboards across clouds, regions, and platforms.

Databricks Marketplace has 1,000+ offerings from 100+ providers spanning industries, enabling customers to discover more than just data, evaluate products faster, and unlock innovation to advance their organization's data and AI initiatives.

Clean Rooms

To combat one of the biggest deterrents to data sharing, a data clean room allows the sharing of data with trusted partners in an isolated environment where no party has direct access or visibility into the other party's raw data or in some cases, proprietary models.

Databricks Clean Rooms provide a secure and privacy-safe environment for enterprises to collaborate with stakeholders on any cloud. Within the clean room, data teams gain the flexibility, scale, and ease of centralized real-time collaboration with their choice of language and workload, pre-defined clean room templates, and no data replication or movement required.

Secure your data with confidence

Security is paramount for enterprises as it forms the bedrock for safe and compliant sharing of sensitive data, ensuring protection and regulatory adherence at every level of operation.

TRUST: Complete transparency into how we secure the platform through penetration testing, vulnerability management and secure software development.

SECURITY: Protect your data and workloads with encryption, network controls, auditing, identity integration, access controls and data governance.

COMPLIANCE: Meet your compliance and regulatory requirements with controls that adhere to certifications and attestations of highly regulated industries.

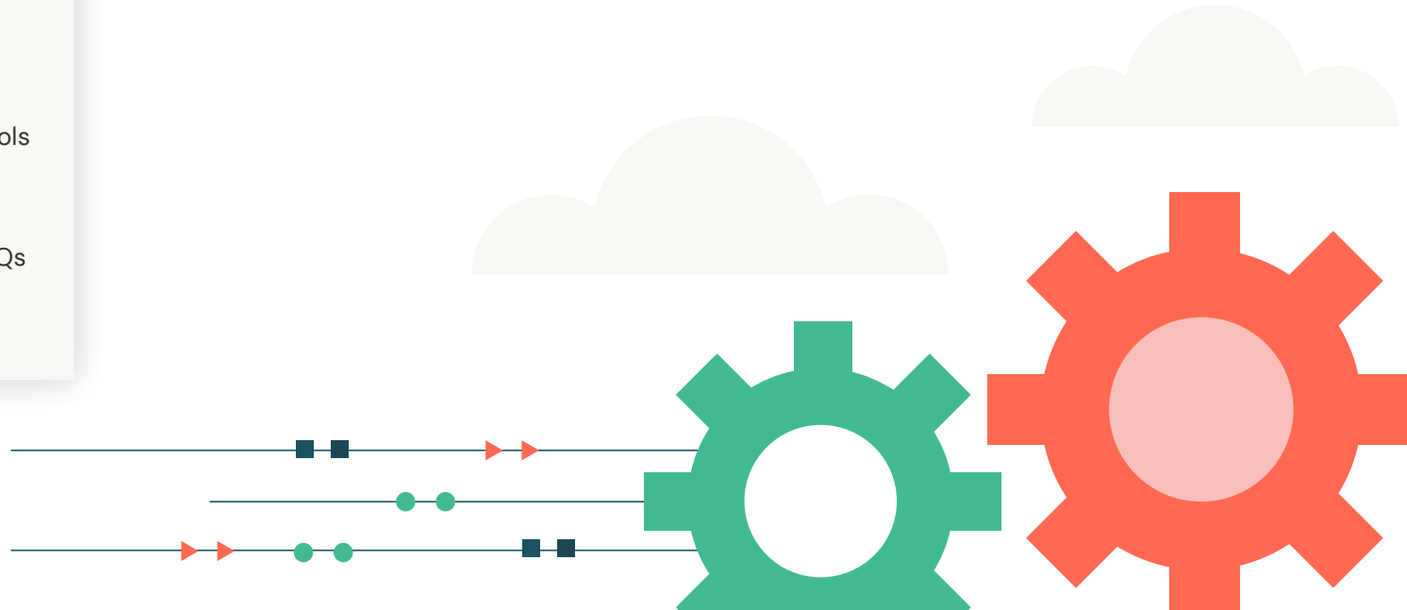
PRIVACY: Adhere to privacy laws and regulations with our prepared Privacy FAQs and documents that transparently set forth how Databricks approaches privacy.

Collaboration is only as strong as your data governance and security

Data teams need to securely discover, access, and work together on trusted data and AI assets. For streamlined governance, enterprises can rely on Unity Catalog, providing multicloud governance for all data and AI assets, including files, tables, dashboards, and machine learning models.

Quickly find, understand, and reference data from across your data estate, boosting productivity. Data search in Unity Catalog is secure by default, limiting search results based on access privileges of the users and adding an additional layer of security for privacy considerations.

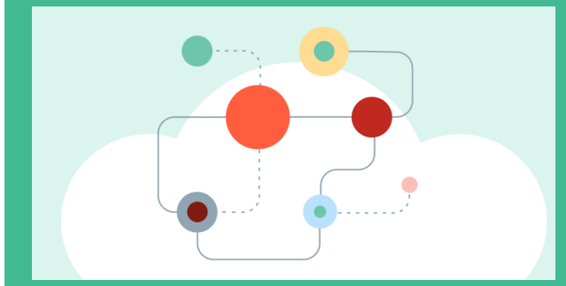
Future proof your data and AI governance with the flexibility to leverage your existing data catalogs and governance solutions like Collibra, Alation, Immuta, and Privacera.



databricks

Industry-specific data sharing use cases

From healthcare and finance to manufacturing and communications, the ability to efficiently and securely share data is not merely a necessity; it's driving innovation. We'll dive into the pivotal role data sharing plays in unleashing the full potential of organization initiatives and driving forward the industries of tomorrow.



Databricks Marketplace Partners

Throughout this eBook, we will be sharing links to relevant Databricks Marketplace and their listings. They can offer instant access and enable data sharing for faster time-to-insights across common and high-impact use cases.

Check them out to learn more and share with your teams to save hours of discovery, design, development and testing — go from idea to proof of concept (POC) within minutes.

CHAPTER 3: INDUSTRY-SPECIFIC DATA SHARING USE CASES

3.1 Healthcare and Life Sciences

Data will power insights that can shift care from reactive to predictive. Across healthcare and life sciences, there has never been more data available than today. In fact, healthcare generates 30% of the world's data volume, and is growing faster annually than any other industry at 36% CAGR.

As more data is generated in the industry, healthcare providers, payers, life sciences, and health tech organizations rely on connecting and collaborating with a growing ecosystem of data partners.

The challenge? Data resides in silos across numerous source systems and point solutions (i.e. electronic health records, laboratory information systems, HR systems, etc.), many of which lack support for unifying both structured and unstructured data. In a highly regulated industry, a lack of unified governance that secures sensitive protected health information (PHI) limits sharing and collaboration, driving up operational cost and complexity as data teams spend cycles securing and governing data.

Common use cases

ACCELERATE DRUG DISCOVERY: by bringing together diverse datasets across teams and organizations, teams are identifying patterns, uncovering potential drug targets, and optimizing therapeutic approaches for patients.

PRESCRIBE NEXT-BEST ACTION: from nudging a patient in the right direction, to arming life sciences sales and marketing with more effective next steps with healthcare professionals, data sharing is helping teams identify what to offer next, at a personalized level.

DRIVE EFFICIENT CLINICAL TRIALS: accessing cross-ecosystem data, including real-world data (RWD), is increasing efficiency in site selection, cohort design, and trial feasibility. Moreover, providers can connect patients with the most promising clinical trials faster.

Data sharing in healthcare and life sciences

When healthcare providers can unlock and share sensitive data internally, externally, and compliantly, organizations can increase clinician productivity, enhance patient personalization, and deploy resilient, digitally-enabled supply chains to respond to varying patient demand. Life sciences organizations are able to increase research and development velocity, personalize next-best action for commercial effectiveness, and streamline manufacturing operations.

Customer spotlight: delivering better patient outcomes by unlocking collaboration

- Kythera Labs builds data solutions to empower the healthcare and life sciences industries, and it is using lakehouse architecture to store and process more than three petabytes of raw healthcare data. Despite the complexity of the data, on Databricks, Kythera Labs accelerated time-to-market by 80% — from two days to two minutes.

- CareSource, a leading Medicaid managed care plan serving over 2 million members, relies on Databricks to provide a unified view into healthcare and member data. With this data, the organization can drive real impact — from accurate invoicing to predictive modeling for proactive healthcare to ensure its millions of members receive optimized and tailored care.

- Amgen uses Databricks to provide a holistic view of its data to streamline operations and accelerate drug discovery. Amgen shares data and insights across the organization securely with deep data governance and granular access controls.

Databricks Marketplace partners

- IQVIA is a leading global provider of advanced analytics, technology solutions, and clinical research services to the life sciences industry dedicated to delivering actionable insights. IQVIA has the world's largest global healthcare data network, with over 1 million data sources providing access to 800M+ non-identified patient records and 95B+ healthcare records processed annually in over 100 countries.

- Ontada is an oncology data science and technology business sharing real oncology de-identified patient-level datasets aligned with NCCN guidelines. Leveraging structured data from more than 2.4 million patients across 85 cancer types, Ontada helps healthcare organizations drive data across diagnosis, treatment, outcomes, and more.

- Datavant is reducing the friction of data sharing across the healthcare industry by building technology that protects the privacy of patients while supporting the linkage of patient health records across datasets. With Datavant, customers can tokenize and de-identify their data natively on Databricks, without moving the data.

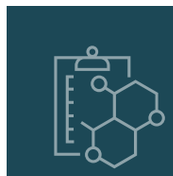
> Delta Sharing is changing how our customers interact and leverage IQVIA data to get more value with less effort. In collaboration with Databricks, life sciences organizations can now access the unparalleled value of IQVIA data in the Databricks Marketplace, empowering our customers to gain invaluable insights for informed and impactful transformations.
>
> **BILL ZANINE**
> Head of U.S. Data Management, Channel and Specialty
> IQVIA

- John Snow Labs, an AI and NLP for healthcare company, provides state-of-the-art software, models, and data to help healthcare and life science organizations build, deploy, and operate AI projects. John Snow Labs provides over 100 data sharing assets on Marketplace.
- Definitive Healthcare transforms data, analytics, and expertise into healthcare commercial intelligence. Through their platform, healthcare practitioners can access NRx and TRx datasets derived from prescription drug claims data from Definitive Healthcare's Rx clearinghouse partners, giving them complete visibility into new and total prescription activity.

Deliver the future of care by equipping your organization with timely and accurate patient insights, medical research results, and device performance to optimize decision-making in critical healthcare and life sciences settings.

Learn how Delta Sharing promotes open data sharing

GET STARTED

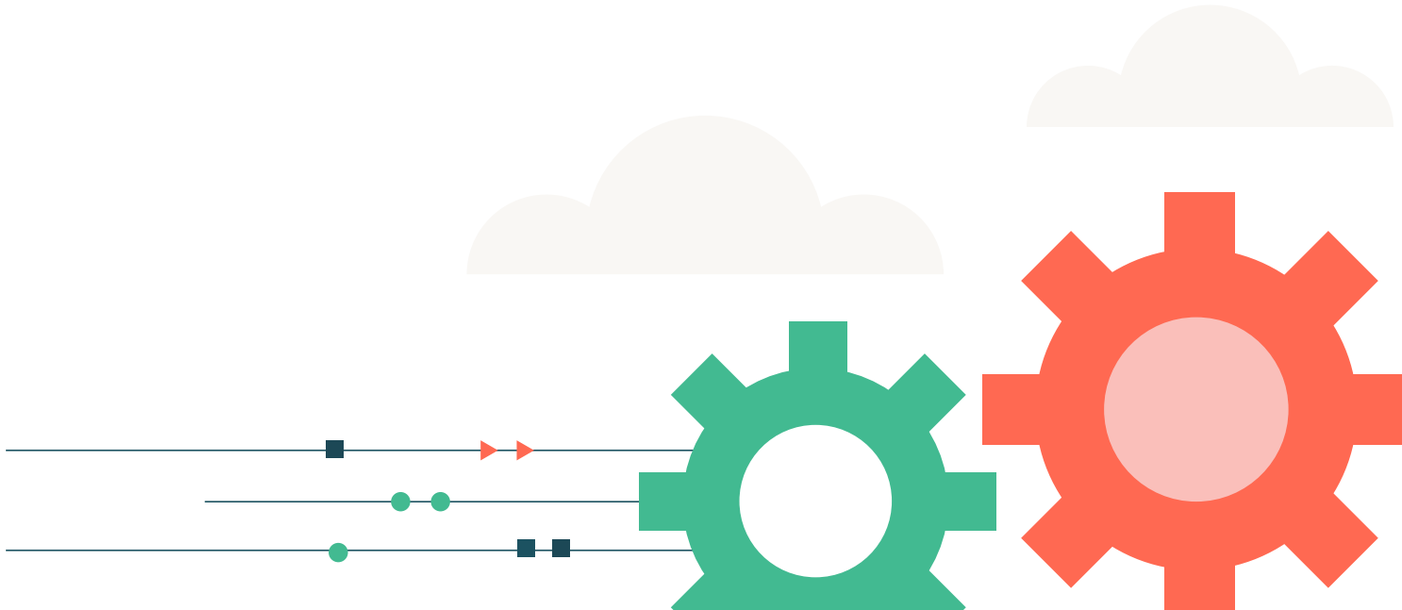


Explore the Marketplace for Healthcare and Life Sciences

BROWSE MARKETPLACE LISTINGS

Try the Databricks Data Intelligence Platform for Healthcare and Life Sciences

GET STARTED



databricks

CHAPTER 3: INDUSTRY-SPECIFIC DATA SHARING USE CASES

3.2 Financial Services

Corporate and retail banking, capital markets, retailers, insurance providers, and the fintech and insuretech companies that power the industry are all custodians of massive amounts of sensitive and complex data. Due to the decentralized nature of these institutions and evolving global regulations, maintaining data quality, creating comprehensive data catalogs, and promoting cross-functional collaboration on legacy systems is challenging. This leads to fragmented governance, divergent definitions, standards, and data comprehension, hindering data's effective use in decision-making.

Data sharing for financial services

Secure open banking and Insurance demand democratizing customer data with customer consent, enabling FSIs to broaden their ecosystems and enter new markets. The Databricks Data Intelligence Platform for Financial Services gives FSIs a simple, unified, governed approach to risk management and compliance, real-time insights, privacy-safe clean rooms for collaboration, and open data sharing and monetization in Databricks Marketplace. To mitigate the data risks associated with SaaS LLMs, FSIs can privately train models within Databricks to maintain data privacy and accuracy.

Common use cases

ACTIVATE OPEN FINANCE: to remain competitive, banks must offer an engaging experience that goes beyond traditional banking via personalized insights, recommendations, setting financial goals and reporting capabilities — all powered by advanced analytics. Open finance is based on data sharing principles that can empower banks to offer a broader range of possibilities to their clients that are suited specifically to their needs.

ANALYZE STOCK MARKET MOVEMENTS: data exploration is the process of investigating a new data set by asking basic questions in order to gain insights for further analysis. Delivering seamless access to vast digital asset data (i.e. blockchain networks, crypto markets, and decentralized finance) is essential to support trading and risk strategies.

MONETIZE ALTERNATIVE DATA: leverage alternative data to analyze or forecast portfolio companies and generate returns. Make better investment decisions from alternative data by uncovering valuable insights about trends, behaviors and risks, including back-testing, market risk and ESG investing.

Facilitating secure financial data sharing

- Block, formerly Square, experienced a 12x reduction in compute cost and 20% lower data egress costs after integrating Unity Catalog for visibility into business units and simple data control. The operational efficiency of data sharing improved significantly, reducing the time required from days to seconds.

- Nasdaq made its data discoverable on the Databricks Marketplace and accessible via Delta Sharing to give customers open access using their choice of tools. It's streamlined data delivery for large data sets, enabling Nasdaq clients to bring their own compute environment. Now, clients can read freshly curated data with little to no integration while Nasdaq continues expanding its high-quality data products.

- London Stock Exchange Group (LSEG) uses Databricks to help customers drive faster time to insights with real-time unified access and enhanced analytics for quantitative decision-making capabilities.

- Jefferies leverages Delta Sharing to enhance data delivery and unlock the business value of data using enriched location data. It also uses Safegraph's location data to power custom insights in financial services.

- SafeGraph uses Databricks' flexibility to streamline customer access and securely reach a broader user base regardless of cloud or platform. Through lakehouse architecture, it has accelerated spatial querying by up to 10x — helping organizations optimize workflows and facilitate the exchange of hard-to-attain geospatial data sets and insights.

> With Nasdaq data now discoverable on the Databricks Marketplace and accessible via Delta Sharing, we will enable customers to access data in an open way across their tools of choice and accelerate their time.
>
> **BILL DAGUE**
> Vice President and Head of Data Product at Nasdaq

Databricks Marketplace partners

- London Stock Exchange Group (LSEG) uses Refinitiv Tick History to provide access to historical financial market data from over 500 global trading venues and third-party contributors. It includes detailed information about trades, such as each transaction's time, price, and volume — cleaned, normalized, and ready for use in research and analysis of trading strategies.
- S&P Global Market Intelligence offers the S&P Global Marketplace Workbench, a web-based notebook environment that allows organizations to test and experiment with datasets from S&P Global and curated third-party providers.
- Nasdaq Data Link is one of the world's most diverse sources of market intelligence from Nasdaq, Inc. connecting asset owners, managers, and all segments of the investing public with a suite of financial, fund, economic, and alternative data, in addition to providing end-to-end data management services and delivery APIs for real-time and delayed data.
- CoreLogic provides information intelligence to identify and manage growth opportunities, improve business performance, and manage risk. Through Databricks Marketplace, all its data assets are discoverable via Delta Sharing — unlocking valuable property-level insights.

Leveraging open data sharing on lakehouse architecture, FSIs can power innovations that drive sustainable value creation at scale. Take advantage of pre-built, financial services-specific offerings to accelerate data-driven transformations on the unified Databricks Data Intelligence Platform.

Learn how Delta Sharing promotes open data sharing

GET STARTED

Explore the Marketplace for Financial Services

BROWSE MARKETPLACE LISTINGS

Try the Databricks Data Intelligence Platform for Financial Services

GET STARTED

CHAPTER 3: INDUSTRY-SPECIFIC DATA SHARING USE CASES

3.3 Retail and Consumer Goods

Retailers, consumer goods companies, bottlers, and distributors must navigate siloed customer data platforms, marketing clouds, and inventory data, but often only come up with a fragmented view of customers. The disparate infrastructures required for these sources slow data processing so much that enterprises cannot deliver on the real-time use cases required to accommodate evolving customer needs and desires. Lacking the technology necessary for efficient data sharing, retail and consumer goods companies are missing out on end-to-end insights, leading to supply chain disruptions, narrowing revenue margins, and costly "trial and error" decision-making.

Data sharing in retail and consumer goods

The Databricks Data Intelligence Platform for Retail provides a secure, connected foundation where retail and consumer goods organizations have governance and security over personal customer data and end-to-end visibility for intentional interventions. To power competitive growth strategies, Databricks Marketplace and Delta Sharing make it easy to quickly integrate new data into existing environments to tap into customer preferences, product performance, and market trends.

Leveraging new models, applications, and data sharing via secure data clean rooms, retail and consumer goods businesses can scale data and AI efficiency, speed, and performance while lowering TCO and driving revenue.

Common use cases

INTEGRATE POINT-OF-SALE (POS) AND EXTERNAL CAUSAL DATA TO INCREASE MARGINS: determine product-level units sold and sales amounts by store location, and calculate real-time inventories across multiple store locations.

GAIN VISIBILITY INTO RETAIL SUPPLY CHAIN TO IMPROVE AVAILABILITY: identify DC fill rates, inventory and to-store shipments as well as store inventories.

UNDERSTAND PURCHASING BEHAVIORS AND ACTIVATE CONSUMERS: identify and understand shopper preferences, personalize their experiences, and promote conversion.

Deliver more value throughout the customer journey

Al-Futtaim (distributor and operator of global brands like Toyota, IKEA, Ace Hardware and Marks & Spencer) uses Databricks to unify and collaborate on previously inaccessible data from more than 200 of its brands. Streamlining its data and AI approach accelerated time-to-insights 4x, decreased lost sale opportunity and inventory holding cost by 30%, and boosted automotive sales by 10%. Now that data teams can integrate point-of-sale (POS) systems with other applications, IoT-enabled devices, and external causal data, organizations can accurately forecast demand, manage inventory, and reduce waste for profitability.

Databricks Marketplace partners

Habu delivers a purpose-built clean room platform for multicloud, multi-party collaboration that enables enterprises to unlock new insights and accelerate growth while maintaining the highest privacy and security standards.

Crisp connects to over 40 U.S. retailers and distributors with real-time point of sale and supply chain data to enable faster and holistic decision-making across all organizational functions: IT, sales, supply chain, marketing, finance, and more.

AccuWeather powers supply chain optimization via real-time feeds of historical and forecast weather data, hyper-localized forecasts, weather targeting, and human insights to improve advertising relevance.

With unprecedented access to these data sources on a modern data and AI platform, retail and consumer goods organizations can act quickly, precisely, and effectively to engage customers for measured revenue growth. Pull the right levers across your value chain to achieve business outcomes from the top down.

Learn how Delta Sharing promotes open data sharing

GET STARTED

Explore the Marketplace for Retail

BROWSE MARKETPLACE LISTINGS

Try the Databricks Data Intelligence Platform for Retail

GET STARTED

> Collaboration has increased between data divisions because it's much easier to share customer data generated by any of our brands to create a golden record of our customer.
>
> **DMITRIY DOVGAN**
> Head of Data Science at Al-Futtaim Group

CHAPTER 3: INDUSTRY-SPECIFIC DATA SHARING USE CASES

3.4 Communications, Media, and Entertainment

The communications, media & entertainment industry is undergoing one of the most significant periods of growth (and change) in its 100+ year history. With more than 4B people on the planet connected to a smartphone, computer, or television; we are seeing an unprecedented arms race with media and entertainment organizations increasingly competing to capture a share of an individual's time and wallet. This increased competition has come at the cost of complexity. A consumer's digital life is global, spread across multiple devices and online environments. And with no one organization able to ringfence the breadth of a consumer's digital sandbox, teams are often left with a fragmented view of their audience. This fragmentation makes it technologically complex and expensive to acquire, engage, and retain customers over time.

As the industry of unstructured data (video, images, and audio workloads), communications, media & entertainment organizations need to consider how their enterprise data platform helps them break down data silos, increase cross-group collaboration, and connect valuable insights to action.

> Leveraging the powerful capabilities of Delta Sharing from Databricks enables Pumpjack Dataworks to have a faster onboarding experience, removing the need for exporting, importing, and remodeling of data, which brings immediate value to our clients. Faster results yield greater commercial opportunity for our clients and their partners.
>
> **COREY ZWART**
> Chief Technology Officer at Pumpjack Dataworks

Data sharing in communications, media, and entertainment

Holistic data management on the Databricks Data Intelligence Platform for Media and Entertainment supports seamless, governed, and secure collaboration for internal and external stakeholders to discover, assess, and share structured, semi-structured, and unstructured data, AI models, notebooks, and dashboards. By engaging with data ecosystem partners, accelerate initiatives like find your audience using overlap analysis, lift analysis, look-a-like modeling, and reach and frequency analysis in privacy-safe clean rooms. Understand your audience through data analysis that fulfills use cases for segmentation, cohort analysis, churn analysis, customer lifetime value, and sentiment analysis. And activate your audience with real-time, personalized content recommendations and ad optimizations based on previous behavior.

Common use cases

FIND THE RIGHT AUDIENCE: identify consumers and engage with hyper-relevant content that satisfies their immediate needs using accurate search and response, keyword mapping, and targeted automatic recommendations.

UNDERSTAND YOUR AUDIENCE BETTER: create advanced customer segments to drive better purchasing predictions based on behaviors using sales data, campaigns and promotions systems.

ACTIVATE YOUR AUDIENCE MORE EFFECTIVELY: understand the behavior of consumers, capture the behavior of various households, build useful customer clusters, and target with the right promotions and offers.

Personalize viewer experiences and boost overall engagement

- Condé Nast realized $6 million in infrastructure cost savings and a 50% reduction in IT operational costs after faster ETL pipelines on Databricks reduced processing times by 60%.
- Kantar used Databricks to accelerate insights from several months to just two days and scaled data processing to analyze 3 million receipts monthly to understand purchase behaviors within defined populations.
- Pumpjack Dataworks manages, secures, and monetizes sports fan data with Databricks to support sports rights holders with insights for targeted experiences, new sponsorship, and revenue opportunities.
- Acxiom relies on Databricks to securely unify customer data to provide its clients with the means to create customer-centric insights from the data points that matter. Now, clients can effectively improve the customer experience and their ability to increase marketing conversions and revenue.

Databricks Marketplace partners

- LiveRamp offers identity resolution solutions in Databricks Clean Rooms to connect disparate offline and online consumer touchpoints to a single identifier for new customer and prospect insights, advertising optimization, and audience segmentation and reach.
- Habu delivers a purpose-built clean room platform for multicloud, multi-party collaboration that enables enterprises to unlock new insights and accelerate growth while maintaining the highest privacy and security standards.
- ActionIQ provides a Customer Data Platform (CDP) that enables customers to integrate and manage customer data. Through Databricks, it grants secured business access to the user experiences necessary for driving personalized engagement.

Maximizing media data efficiency is essential to delivering the experiences expected in modern communications, media, and entertainment. Databricks provides the platform necessary for efficiency, with the data sharing capabilities required for applications in the fast-paced CME industry.

Learn how Delta Sharing promotes open data sharing

GET STARTED

Explore the Marketplace for Media and Entertainment

BROWSE MARKETPLACE LISTINGS

Explore the Marketplace for Advertising and Marketing

BROWSE MARKETPLACE LISTINGS

Try the Databricks Data Intelligence Platform for Media and Entertainment

GET STARTED

CHAPTER 3: INDUSTRY-SPECIFIC DATA SHARING USE CASES

3.5 Manufacturing and Energy

The manufacturing and energy industry relies on complex, unstructured, and distributed industrial data sets to produce automobiles, electronics, chemicals, food and beverages, utilities, renewable energy, oil and gas, and aerospace engineering. Interoperability between software, hardware, and devices is impossible among legacy data and organizational silos, strict data sharing regulations, and cultural resistance to change. Without reliable real-time data across production processes, manufacturing and energy enterprises struggle to extract and analyze data cost-effectively for informed decision-making.

> We recognize that openness of data will play a key role in achieving Shell's Carbon Net Zero ambitions. Delta Sharing provides Shell with a standard, controlled, and secure protocol for sharing vast amounts of data easily with our partners to work toward these goals without requiring our partners be on the same data sharing platform.
>
> **BRYCE BARTMANN**
> Chief Digital Technology Advisor, Shell

Data sharing in manufacturing

The Databricks Data Intelligence Platform for Manufacturing unifies data, analytics, and AI workloads with secure live sharing and built-in data governance to deliver compliant access to clean data and analytics on a single platform. Purpose-built solution accelerators guide enterprises through a variety of data science best practices including LLM model builds, supply chain optimization, defect detection solutions and more. These fully functional notebooks and tools accelerate project results, return on investment and reduce time on development and testing. With a robust data foundation and safe channels for data sharing, manufacturers have the visibility necessary to improve operations downstream based on what business leaders need upstream.

For example, real-time production and process data from machine settings and operating conditions highlight opportunities to decrease waste, reallocate resources, and enable predictive maintenance to prevent business disruptions and lower recurring costs. Sharing data on supplier performance, inventory levels, production schedules, and logistics supports supply chain planning and optimization. Moreover, collaborative product development enhances design and production among disparate factories and owners to speed time-to-market and improve quality.

Common use cases

OPTIMIZING & DIGITIZING THE SUPPLY CHAIN: leverage data from connected factory equipment, order flows, supplier performance, and production and logistics execution to power real-time insight into operational performance that improves quality, on-time delivery and lead times across the supply chain.

IMPLEMENTING DIGITAL TWINS: manufacturers have turned to building digital twins, which are virtual representations of objects, products, pieces of equipment, people, processes, or even complete manufacturing ecosystems.

FOSTERING PREDICTIVE MAINTENANCE: knowing when equipment needs maintenance is critical to companies - unplanned downtime is a major disruption to our businesses costing millions of dollars. An important way to save time and money is to use machine learning to better predict outages earlier and plan maintenance work before the failure occurs.

Driving impact across the entire value chain

- Shell is utilizing Databricks' centralized, controlled, and secure data-sharing protocols at scale to achieve its Carbon Net Zero ambitions. Being able to share massive amounts of data with partners, Shell can work toward its goals without requiring partners to be on the same data sharing platform.

- Kotahi is delivering on its vision of a sustainable, efficient supply chain for the future with the help of Databricks. Through the Databricks platform, they are able to share valuable data insights to help reduce waste in the supply chain and streamline container movements.

Databricks Marketplace partners

- Rearc Container Port Traffic delivers a purpose-built clean room platform for multicloud, multi-party collaboration that enables enterprises to unlock new insights and accelerate growth while maintaining the highest privacy and security standards.
- S&P Global Market Intelligence offers the S&P Global Marketplace Workbench, a web-based notebook environment that allows organizations to test and experiment with datasets from S&P Global and curated third-party providers.
- PredictHQ offers analytics for manufacturing in the Databricks Marketplace that can lower the barrier to data-driven innovation. Manufacturers can deliver more robust reporting, analytics, and predictions in their business by unifying contextualized time series data from industrial assets with a vast array of enterprise data sources.

Embrace the benefits of applying data, analytics, and AI on a platform that supports simplicity and integration for unobstructed insights across complicated manufacturing and energy value chains. Not only does the Databricks Data Intelligence Platform for Manufacturing lower costs and boost productivity, but it also lays a foundation to scale for the future.

Learn how Delta Sharing promotes open data sharing

GET STARTED

Explore the Marketplace for Manufacturing

BROWSE MARKETPLACE LISTINGS

Explore the Marketplace for Energy

BROWSE MARKETPLACE LISTINGS

Try the Databricks Data Intelligence Platform for Manufacturing

GET STARTED

CHAPTER 3: INDUSTRY-SPECIFIC DATA SHARING USE CASES

3.6 Public Sector

Federal, state, municipal, and local governments, as well as higher education and research institutions generate a wealth of highly sensitive unstructured data that must be governed and strictly regulated. Concerns about exposing sensitive citizen and national security information demand data traceability, but legacy technology, data silos and lack of real-time capabilities limit what these organizations can accomplish without potentially compromising security.

Governments need to increase the speed of service delivery and data sharing across departments, between agencies and with private companies to enhance the constituent experience and improve outcomes for citizens.

Data sharing in public sector

Delta Sharing's native integration with Unity Catalog on the Databricks Data Intelligence Platform for Public Sector lets organizations manage, govern, audit, and track shared data usage. Organizations gain an added layer of security and efficiency to conserve costs on a single, collaborative platform. Databricks cloud services have FedRAMP and DoD authorizations to operate (ATOs) across various clouds and regions, and support public sector mission objectives and the data protection requirements of government agencies.

Furthermore, the solution accelerators on the Databricks Data Intelligence Platform for Public Sector are custom-built to fuel fast, reliable, and applicable innovations to keep up with the volume and velocity of its data.

With the emergence of SaaS LLMs and the data leaks they've led to, regulated industries must privately train models to control data privacy and hallucination. The Databricks Platform provides a safe environment for organizations to fulfill various use cases with Generative AI, including LLMs for open source intelligence, predictive maintenance to reduce fleet downtime and costs, and constituent 360 to improve processes and outcomes across disparate departments to offer citizens a service experience that mirrors those of private companies.

Common use cases

FRAUD DETECTION: data sharing can aid in the detection and prevention of fraud and identity theft. By cross-referencing data from various agencies, it becomes easier to spot inconsistencies or suspicious activities associated with everything from tax returns to claims requests.

CONSTITUENT 360: with access to a comprehensive profile of each citizen, public sector agencies can tailor their services and communication to individual preferences and needs. This personalization can improve the citizen's experience and satisfaction with government services.

CRISIS RESPONSE AND MANAGEMENT: during natural disasters, pandemics, or other crises, sharing data among government agencies, first responders, and healthcare organizations can enhance situational awareness, resource allocation, and coordination. This includes sharing real-time information on weather, emergency response efforts, and healthcare resources.

Delivering better public services

- United States Postal Service Office of the Inspector General (USPS OIG) uses Databricks to ingest and process customer and delivery-related data with a centralized view across teams for downstream analytics and machine learning. This foundation was critical during the 2020 election when voting by mail spiked due to the pandemic. USPS needed to audit, investigate, and uncover anomalous activity in real time, and with the Databricks Platform, it could.

- U.S. Citizenship and Immigration Services (USCIS) leverages the Databricks Platform to federate immigration and related data across dozens of disparate sources. This capability has unlocked operational efficiencies and opportunities for its data organization to drive business intelligence and fuel ML innovations that can streamline the visa application and petition processes.

> With a data team of more than 100 people who needed to work together or respond to anomalous activities in a timely manner, the OIG looked to the cloud and a new data architecture that would offer all its data teams easy access to any data and unlock new analytical and machine learning capabilities to further its efforts to improve mail delivery efficiency and accountability.
>
> Office of Inspector General Customer Story

Databricks Marketplace partners

- Bloomberg Government Data License is one of the largest sources of quality federal government contracting opportunities and market intelligence data, designed to optimize the steps in the federal government contracting lifecycle. With 31+ million government contracts updated daily and over ten years' worth of historical agency spend, the Bloomberg Government Data License provides valuable context to help federal contractors set current strategies and make long-term business decisions.
- FiscalNote is a leading technology provider of global policy and market intelligence. Through the Marketplace, they aim to help the public sector navigate changing political, corporate, and regulatory environments with macroeconomic, news pulse, and intelligent sentiment data.
- Precisely is the global leader in data integrity, providing accuracy and consistency in data for 12,000 customers in more than 100 countries, including 99 of the Fortune 100. Precisely's products for data integration, data quality, and data governance, as well as location intelligence and enrichment, are designed to power better business decisions and create better outcomes.

Historically, the public sector has struggled to keep up with commercial industries. Now armed with the same data innovations and enterprise best practices, agencies can realize similar outcomes. Maximize what's possible with public sector data sharing and secure collaboration to optimize the foundational systems that we all depend on.

Learn how Delta Sharing promotes open data sharing

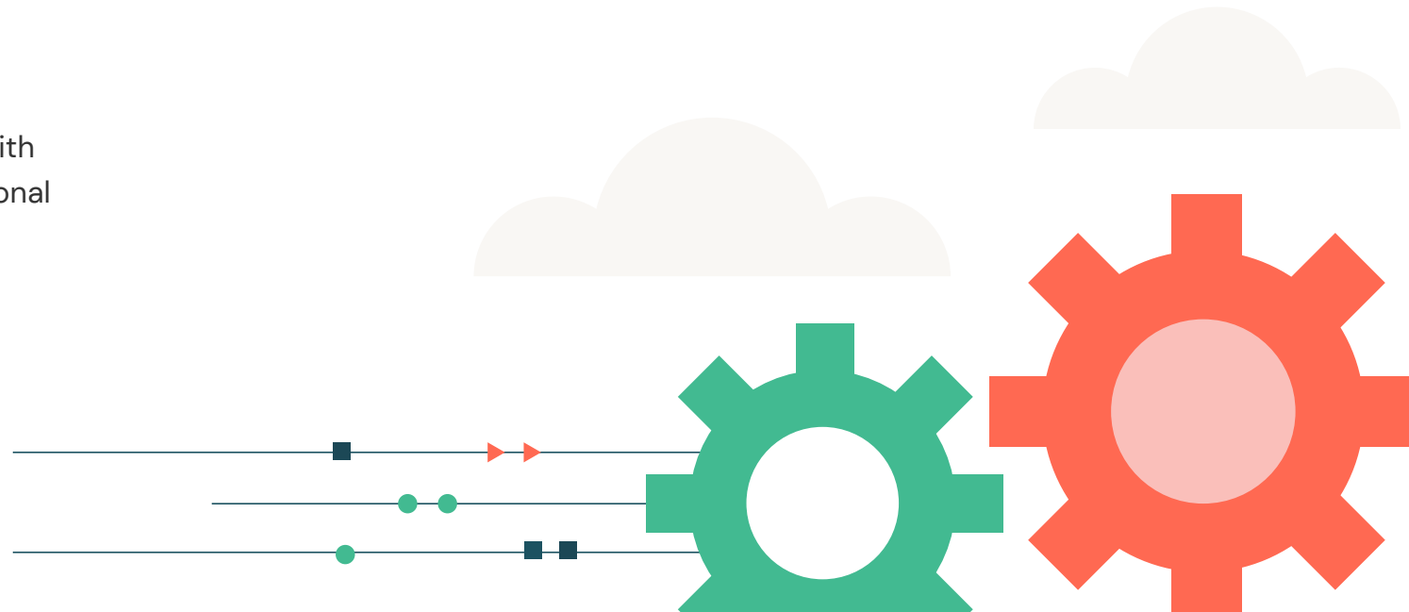
GET STARTED

Explore the Marketplace for Public Sector

BROWSE MARKETPLACE LISTINGS

Try the Databricks Data Intelligence Platform for Public Sector

GET STARTED



databricks

CHAPTER 4

Gaining value through open data sharing

The Databricks Data Intelligence Platform is built for data sharing and collaboration, providing a centralized infrastructure to support your organization's needs both internally and externally with Delta Sharing, Unity Catalog, and Databricks Clean Rooms. The Databricks Marketplace is powered by Delta Sharing and enables consumers to quickly discover, access and evaluate data, analytics and AI assets, as well as securely collaborate on more than just data to accelerate innovation, optimize data science projects and fulfill use cases quickly and efficiently.

About Databricks

Databricks is the data and AI company. More than 10,000 organizations worldwide — including Comcast, Condé Nast, and over 50% of the Fortune 500 — rely on the Databricks Lakehouse Platform to unify their data, analytics and AI. Databricks is headquartered in San Francisco, with offices around the globe. Founded by the original creators of Apache Spark™, Delta Lake and MLflow, Databricks is on a mission to help data teams solve the world's toughest problems. To learn more, follow Databricks on Twitter, LinkedIn and Facebook.

START YOUR FREE TRIAL

Contact us for a personalized demo
databricks.com/contact

databricks