



Hello Databricks rockstars!

It's March 2024! Here is your quick recap of Databricks platform enhancements, key releases, previews, blogs, and solutions from the last month, as well as training and events in the coming month. We highly recommend you follow up with your account team.

## Table of Contents

[The Lakehouse Platform - February 2024 Release Notes](#)

[Data Engineering & Analytics](#)

[Data Science, ML, and Generative AI](#)

[Security, Governance, and Data Asset Collaboration](#)

[Customer Solutions](#)

[Workshops, Webinars, and Events](#)

---

## Product Safe Harbor Statement

*This 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.*

---

**SAVE THE DATE**

# DATA+AI SUMMIT

Databricks launched the world's first data lakehouse in the cloud, and Data + AI Summit is where you can discover how lakehouse brings together all your data so you can perform everything from BI to AI on a single, open platform.

Choose from over 500 breakout sessions and hear from data and AI thought leaders about latest trends and innovations, including LLMs, Apache Spark, Delta Lake, MLflow, TensorFlow, dbt and more.

**16K+** ATTENDEES ON SITE

**500+** BREAKOUT SESSIONS

**15+** TRAINING SESSIONS

**WORLD'S LARGEST DATA, ANALYTICS AND AI CONFERENCE**

**IN-PERSON JUNE 10-13 SAN FRANCISCO**



**Virtual Event // June 12 - 13 // Free live stream of keynotes + select sessions**

**REGISTRATION IS OPEN!**

**In Case You Missed It** – You can still catch our Q1 Product Roadmap Webinar on Demand. [Just register](#) to get the link


## The Lakehouse Platform – February 2024 Release Notes

- **GA**: Unity Catalog Volumes ([AWS](#), [Azure](#), [GCP](#)) ([Blog](#)) ([Demo Video](#)) – Discover, govern, query and share any non-tabular data
- **GA**: File arrival triggers in Databricks Workflows ([AWS](#), [Azure](#), [GCP](#))
- **GA**: Databricks Git server proxy ([AWS](#), [Azure](#), [GCP](#))
- **GA**: Workspace file support for the dbt and SQL file tasks ([AWS](#), [Azure](#), [GCP](#))
- **GA**: Databricks Connect for Scala ([AWS](#), [Azure](#), [GCP](#))
- **GA**: Universal Format (UniForm) for Iceberg compatibility with Delta tables
  - in DBR 14.3 LTS ([AWS](#), [Azure](#), [GCP](#))
- **PUBLIC PREVIEW**: Support for cloud storage firewall from serverless compute ([AWS](#))
- **PUBLIC PREVIEW**: Delta Sharing supports tables that use deletion vectors ([AWS](#), [Azure](#), [GCP](#))
- **PUBLIC PREVIEW**: Native XML file format support in DBR 14.3 LTS ([AWS](#), [Azure](#), [GCP](#))
- **PUBLIC PREVIEW**: Support for Cloudflare R2 storage to avoid cross-region egress fees ([AWS](#), [Azure](#), [GCP](#)) ([Blog](#)) – You can now use Cloudflare R2 as cloud storage for data registered in Unity Catalog. Cloudflare R2 is intended primarily for Delta Sharing use cases in which you want to avoid the data egress fees charged by cloud providers when data crosses regions. R2 storage supports all of the Databricks data and AI assets supported in AWS S3, Azure Data Lake Storage Gen2, and Google Cloud Storage.
- Notebooks for monitoring and managing Delta Sharing egress costs are now available ([AWS](#), [Azure](#), [GCP](#))
- Add data UI supports XML file format ([AWS](#), [Azure](#), [GCP](#))
- Use AI Functions to invoke a generative AI model from Foundation Model APIs ([AWS](#), [Azure](#))
- Full-page AI-powered search ([AWS](#), [Azure](#), [GCP](#))
- Run SQL notebook jobs on a SQL warehouse ([AWS](#), [Azure](#))
- Search for machine learning models in Unity Catalog using global workspace search ([AWS](#), [Azure](#), [GCP](#))
- Databricks Git server proxy no longer requires CAN\_ATTACH\_TO permissions ([AWS](#), [Azure](#), [GCP](#))
- Create tables from files in Unity Catalog Volumes ([AWS](#), [Azure](#), [GCP](#))
- UI experience for OAuth app registration ([GCP](#))
- **GA**: Databricks Runtime 14.3 LTS ([AWS](#), [Azure](#), [GCP](#))
  - [Spark catalog APIs now fully supported in shared access mode](#)
  - [Delta UniForm is now generally available](#)
  - [New SQL function EXECUTE IMMEDIATE](#)

- [Recompute data skipping statistics for Delta tables](#)
- [Query state information for stateful streaming queries](#)
- [Use Microsoft Entra ID \(formerly Azure Active Directory\) for Kafka authentication on shared clusters](#)
- [Support added for file and partition pruning to improve query performance](#)
- [Cloud Fetch is now enabled by default](#)
- [Declare temporary variables in a SQL Session](#)
- [Thriftserver updates to remove unused features](#)
- [Use truststore and keystore files in Unity Catalog volumes](#)
- [Native XML file format support \(Public Preview\)](#)
- [Support for Cloudflare R2 storage \(Public Preview\)](#)
- **Blogs, Solutions, and Partner News**
  - [Furthering our commitment to responsible AI development through industry and government organizations](#) - At Databricks, we've upheld principles of responsible development throughout our long-standing history of building innovative data and AI products. We are committed to blue-sky research and open source innovation – it's part of our culture and stems from our company's academic roots. Building on this legacy, [Databricks recently joined several industry and government efforts to promote innovation and advocate for the use of safe and trustworthy AI](#). Whether it's contributing insights from our Data Intelligence Platform for research or joining other players to educate policymakers on AI, Databricks is proud to be a part of initiatives to accelerate progress, improve safety, bolster security and harness trust in AI.
  - (Video):  Modernizing your government data estate with a lakehouse architectu...
  - (Video):  Advancing Spark - Azure Databricks News Dec 2023 - Jan 2024 - a great way to walk through the release notes with an expert!

## Data Engineering & Analytics

- [Performance Improvements for Stateful Pipelines in Apache Spark Structured Streaming \(Part 1\)](#)
- [A Deep Dive into the Latest Performance Improvements of Stateful Pipelines in Apache Spark Structured Streaming \(Part 2\)](#)
- [Faster Lakeview Dashboards with Materialized Views](#) - accelerate your BI
- We are excited to announce the upcoming general availability of **Azure Private Link support for Databricks SQL (DBSQL) Serverless**, planned in April 2024, **with no additional charges for use**. We are also thrilled to announce that Azure Storage firewall support with stable VNet subnet IDs is now generally available for DBSQL Serverless. This [blog](#) will give an overview of the two features and associated best practices for securely accessing data in your Azure Storage account from Databricks serverless.
- [ai\\_query on DBSQL - Enhancing your data analysis with Gen AI](#)

- [Unleashing the power of Generative AI with Databricks SQL](#) - [AI functions](#) are built-in DB SQL functions allowing you to access Large language Models (LLM) directly from SQL. Previously, data scientists used [Databricks Model Serving](#) to deploy and manage all AI models: e.g. Foundational models (Llama, MPT, Mixtral) or External Model (Open AI, AWS Bedrock, Anthropic, AI21 etc.). Now with the introduction of AI functions, analysts or business users who do not have an ML background can harness the power of these world-class LLMs within the normal flow of writing and running SQL queries and also share results in BI tools. ETL analysts can also call LLMs in data pipelines to enrich the datasets.
- [Ingesting Data from API Sources](#) - Databricks seamlessly integrates with the application and data infrastructure of organizations. Its ability to extract data from various sources, perform transformations, and integrate with data sinks simplifies system integration differentiating Databricks. This stands in contrast to cloud data warehouses that are reliant on external functions for these tasks that increase complexity and cost.
- [What's new in data engineering and streaming](#) - Databricks was recently named a leader in The Forrester Wave™: Cloud Data Pipelines, Q4 2023. Read the blog to see all of our latest innovations in this space - streaming, enhanced flow control in workflows, DLT, Unity Catalog, and more
- [Databricks System Table Deep Dive – Data Tagging Design Patterns](#)
- [Creating Delta Lake Tables in Databricks Unity Catalog with Azure Data Factory](#)
- [Converting chained stored procedures to Databricks](#)
- [Performance of Querying Uniform-Iceberg Tables in Snowflake written by Databricks](#)
- [Top 10 reasons why your cloud data platform migration should leave stored procedures in the past with a perspective on the future](#)
- (Video)  [Data Engineering in the Age of AI: Data Intelligence Platforms](#)

## Data Science, ML, and Generative AI

- [OLMo is Here, Powered by Databricks](#) - Allen Institute for AI (AI2) released [OLMo 7B](#), an open source, state-of-the-art large language model. Databricks is proud to have supported their work: OLMo (short for Open-source Large Language Model) was trained using our [Mosaic AI Model Training Platform](#). The AI2 team is also sharing the pre-training data and training code used to develop this model (which is a derivative of the [MosaicML LLM Foundry](#)).
- [Fine-Grained Human Feedback - Better Rewards for Language Model Training](#)
- [The Unconscious Patient Problem: A Look at the Importance Of Entity Resolution in Healthcare and Life Sciences](#) - Partner blog featuring Stardog
- [Linking the unlinkables; simple, automated, scalable data linking with Databricks ARC](#) - The addition of automated data linking to [ARC](#) is an important contribution to the realm of entity resolution and data integration. By connecting datasets without a common key, enterprises can harness the true power of their data, drive internal innovation and modernisation, and better serve their customers/clients/citizens/patients/...
- [The End of third-party cookies - Revolutionizing Tech Marketing and The Synergy of PyMC and Databricks](#)

- [LIMIT - Less is More for Instruction Tuning](#) - advise for effectively and cheaply “instruction finetune” an LLM
- [The Shift from Models to Compound AI Systems](#) - AI caught everyone’s attention in 2023 with Large Language Models (LLMs) that can be instructed to perform general tasks, such as translation or coding, just by prompting. This naturally led to an intense focus on models as the primary ingredient in AI application development, with everyone wondering what capabilities new LLMs will bring. As more developers begin to build using LLMs, however, we believe that this focus is rapidly changing: **state-of-the-art AI results are increasingly obtained by compound systems with multiple components, not just monolithic models.**
- (Video) - [AI Model Sharing with Delta Sharing](#)
- (Video) - [Advancing AI - Databricks Vector Search Index](#)

## Security, Governance, and Data Asset Collaboration

- [Strengthening Cyber Resilience through Efficient Data Management: A Response to M-21-31](#) - Securely store and share data across public sector agencies and maintain regulatory compliance with Deloitte and Databricks - featuring [Deloitte's M-21-31 Cybersecurity solution](#) on the the Databricks Data Intelligence Platform
- [An Ultimate Guide to Databricks Unity Catalog](#)
- (Video) - [Azure Databricks Networking Security \(Part 1\) - 2024.02.02](#)

## Customer Solutions

Accelerate onboarding new use cases and your path to production with our pre-packaged Databricks Solution Accelerators and partner BrickBuilder solutions.

- Consulting and SI Partner Offerings: [Brickbuilder Solutions and Accelerators](#)
  - [Databricks adds new migration Brickbuilder Solutions to help customers succeed with AI](#) - See new Brickbuilder migration Solutions from Blueprint, Cognizant, Datapao, Deloitte, Hitachi, Insight, Perficient, Tech Mahindra
  - **Adastra:** [Enabling an Enterprise Modern Data Warehousing with Azure Databricks](#)
  - **Avanade:** [Delta Lake Database/Schema Maintenance](#)
  - **Avanade:** [Connect Databricks using ODBC/JDBC with Azure Service Principal's Bearer Token](#)
  - **Blueprint:** [Brickbuilder Accelerator: Unity Catalog Migration](#)
  - **Blueprint:** [The Business of Databricks: The Case for Consolidation](#)
  - **Blueprint:** [The Lakehouse Optimizer](#)
  - **Tiger Analytics:** [A comprehensive guide: Optimizing azure databricks operations with unity catalog](#)
  - **Tredence:** [How Databricks and MLflow could accelerate LLM and GenAI competency in data science](#)
- Databricks Solution Accelerators ([complete](#) list of offerings - and checkout the [github repos](#))

- Learn Databricks with dbdemos.ai ([Video](#) from Advancing Spark)
- Databricks Labs Projects, checkout [GitHub](#)!
- Customers in the News
  - **Antuit.ai** - [Accelerating Success with Databricks: A Deep Dive into antuit.ai's Decision and Customer Impact](#) - In the dynamic realm of AI-driven forecasting, businesses navigate a landscape where strategic choices shape their trajectory. One such pivotal decision was made by a leader in AI-powered forecasting solutions- [antuit.ai](#), now part of [Zebra Technologies](#). Antuit.ai chose to anchor their endeavors on a lakehouse architecture with the Databricks Data Intelligence Platform rather than treading the path of in-house infrastructure
  - **Hanwaha Corporation**: [Powering efficient manufacturing with Databricks](#). "Databricks has helped us improve our work productivity by streamlining data collection, analysis and visualization — creating the foundational standard for data-driven decision-making company-wide." — Haejin Jeong, Chief Digital Transformation Officer, Hanwaha Corporation
  - **Jobis & Villains** - [Transforming tax services with Databricks](#). "Databricks has played a key role in shaping our company's data-driven culture — empowering our entire organization to make faster and more accurate decisions to scale our business." —Jungwoo Hur, COO, Jobis & Villains
  - **M Science** - [Turning alternative data into actionable insights](#).
  - **Netacea** (Video): [Netacea Enables Defensive AI with Databricks](#). Databricks powers Netacea's ability to stop bot attacks at scale with defensive AI. With Databricks, the firm can now ingest up to 150K requests per second, a 50x increase from its legacy system capabilities.
  - **OTTO FUCHS**: [From data to value - revolutionizing a metal-casting process with Digital Twins and the Lakehouse Platform](#) "For us, it was soon clear that only the Databricks Lakehouse architecture, a solution that combines the best of data lakes and data warehouses could meet our needs. The Lakehouse architecture is scalable across various use-cases, from SQL analytics to machine learning and works for numerous of our use-cases."
  - **Specsavers(Video)**: [Specsavers provides the best value optometry and audiology with Databricks](#). Specsavers is on a mission to improve the lives of its customers through hearing and eye care. With the help of Databricks, Specsavers can better personalize their products and more thoroughly understand customer needs to provide the best optometry and audiology care in Europe.
  - **Tabcorp** (Video): [Tabcorp provides personalized gaming experience with Databricks](#). Tabcorp is using data and AI to provide a personalized gaming and entertainment experience for its customers, whether they are watching at home or attending a live event. By unifying data on Databricks, Tabcorp is decommissioning legacy platforms and investing in more AI initiatives to benefit its customers. Using Unity Catalog for governance gives the Tabcorp team a clear and comprehensive

understanding of its data sources, lineage and quality while allowing for a scalable data governance strategy.

- **Trek:** [Trek Bicycle uses Databricks and Qlik to unify point-of-sales data](#). "How do you scale up analytics without blowing a hole in your technology budget? For us, the clear answer was to run all our workloads on Databricks Data Intelligence Platform and replicate our data in near real-time with Qlik." — Garrett Baltzer, Software Architect, Data Engineering, Trek Bicycle
- **US Air Force:** [US Air Force Hackathon - how LLMs will revolutionize USAF flight test](#)
- **Welldoc:** [Welldoc® and Databricks: Enhancing Cardiometabolic Care with Improved Data for Tailored Interventions](#). Welldoc's digital coaching platform is designed to provide personalized support and guidance to individuals managing cardiometabolic conditions. It combines AI with real-time data to create a dynamic and adaptable digital coaching experience. Welldoc's comprehensive [chronic care platform](#) provides multi-condition support across prediabetes, diabetes, hypertension, heart failure, and weight management, along with integrated mental wellbeing and sleep apnea support, supporting connected "care anywhere" and enabling health plans and health systems to extend face-to-face models.



## Workshops, Webinars, and Events

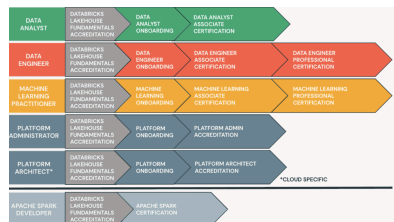
Databricks regularly offers free webinars, workshops, office hours, lunch and learns, and regional events. Here is a quick listing of what's coming up in the next month. Beyond this, you may discover regional events if you join a user group at [Databricks Community](#). You may also find self-paced and instructor-led training and certification pathways at [Databricks Academy](#).

### Free Training and Onboarding Sessions



**Expert Solution Architect led 90-min interactive sessions** 11:00 am Pacific, roughly every 2 weeks. [Register](#) for your favorite topics (Also on-demand)

- Introduction to Delta and Lakehouse (Wednesdays)
- Delta and Databricks SQL (Tuesdays)
- End to End ML on Databricks (Tuesdays)
- Deep Learning on Databricks (Wednesdays)
- Delta in Production (Fridays)
- Delta Live Tables (Mondays)
- Unified Governance with Unity Catalog (Fridays)



Databricks offers free onboarding and advanced training with **live instructors** across regions

#### [Foundational Classes](#)

##### Getting Started With...

- Platform Administration
- Data Engineering on Databricks
- Data Analysis on Databricks
- Machine Learning on Databricks
- Databricks for Business Leaders

#### [Intermediate/Advanced Classes](#)










- Data Pipelines with Delta Live Tables
- Data Analytics and Warehousing with Databricks SQL
- Data Workloads with Repos and Workflows
- Data Governance with Unity Catalog
- Streaming with Structured Streaming and DLT
- Data Storage and Management with Delta Lake
- Performance Optimization with Delta Lake, Apache Spark, and Photon









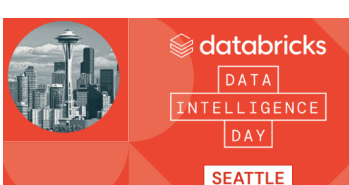


**Databricks Office Hours** - Drop In sessions, LIVE Q&A, Get the Answers to your Questions. [Register](#)

March 20 - 8:00am PST  
March 27 - 10:00am GMT



<div> Webinars and Events </div> <div> (check for updates: <a href="https://www.databricks.com/events">https://www.databricks.com/events</a>) </div>		
<div>  </div> <div> Virtual Event </div> <div> <b>Democratizing the EU's Data</b> </div> <div> Virtual Event  March 05   10:00 AM </div> <div> Join this webinar to hear how Databricks and Delaware are helping shape Europe'... </div> <div> <a href="#">Register now →</a> </div>	<div>  </div> <div> <b>Virtual Retail &amp; Consumer Goods Forum</b> </div> <div> Virtual Event  March 05   09:00 AM </div> <div> The speed of change in retail is only getting faster, and generative AI is... </div> <div> <a href="#">Register Now →</a> </div>	<div>  </div> <div> <b>Financial Services Forum</b> </div> <div> New York, NY  March 06   01:00 PM </div> <div> Join us for the Financial Services Forum to network with industry peers and don'... </div> <div> <a href="#">Register Now →</a> </div>
<div>  </div> <div> Virtual Event </div> <div> <b>Simplify Your Streaming With Delta Live Tables</b> </div> <div> Virtual Event  March 08   12:00 PM </div> <div> Discover how Delta Live Tables simplifies data transformation and ETL. </div> <div> <a href="#">Register now →</a> </div>	<div>  </div> <div> Conference </div> <div> <b>Databricks at HIMSS</b> </div> <div> Orlando, FL  March 11   08:00 AM </div> <div> Join Databricks at HIMSS March 11-15 in Orlando. </div> <div> <a href="#">Event Details →</a> </div>	<div>  </div> <div> Conference </div> <div> <b>Databricks at Gartner Data &amp; Analytics</b> </div> <div> Orlando, FL  March 11   09:00 AM </div> <div> Join Databricks at Gartner Data &amp; Analytics Summit March 11-13 in Orlando. </div> <div> <a href="#">Register Now →</a> </div>
<div>  </div> <div> Meetup </div> <div> <b>Databricks Meetup</b> </div> <div> Austin, TX  March 12   04:00 PM </div> <div> Join Databricks at Easy Tiger South for networking, food and drinks! This is the... </div> <div> <a href="#">Event Details →</a> </div>	<div>  </div> <div> Virtual Event </div> <div> <b>The Gen AI Payoff in 2024</b> </div> <div> Virtual Event  March 14   09:00 AM  March 15   12:00 PM  March 19   09:00 AM </div> <div> Discover the new cost-effective solution to build your own LLM trained on your... </div> <div> <a href="#">Event Details →</a> </div>	<div>  </div> <div> <b>Databricks SQL on the Data Intelligence Platform</b> </div> <div> Virtual Event  March 14   09:00 AM </div> <div> Join this interactive hands-on workshop to learn how Databricks SQL allows you... </div> <div> <a href="#">Register Now →</a> </div>

 <p>Conference</p> <p><b>Databricks at Game Developers Conference</b></p> <p>San Francisco, CA March 18   09:00 AM</p> <p>Join Databricks at GDC March 18–22 in San Francisco.</p> <p><a href="#">Event Details →</a></p>	 <p>Virtual Event</p> <p><b>Data Engineering in the Age of AI</b></p> <p>Virtual Event March 19   09:00 AM</p> <p>This introductory workshop caters to data engineers seeking hands-on...</p> <p><a href="#">Event Details →</a></p>	 <p>Regional Event</p> <p><b>Bangkok Hands-On Workshop</b></p> <p>Bangkok March 20   08:30 AM</p> <p>Shaping the future of data and AI. Discover the Data Intelligence Platform.</p> <p><a href="#">Register now →</a></p>
 <p>Regional Event</p> <p><b>Data Intelligence Day Melbourne</b></p> <p>Melbourne, Australia March 20   08:00 AM</p> <p>Now GenAI is for everyone. Discover the platform that makes it possible.</p> <p><a href="#">Register now →</a></p>	 <p>Virtual Event</p> <p><b>Virtual Government Forum</b></p> <p>Virtual Event March 21   09:00 AM</p> <p>Government agencies have struggled with integrating data from siloed intern...</p> <p><a href="#">Event Details →</a></p>	 <p>Regional Event</p> <p><b>Data Intelligence Day Chicago</b></p> <p>Chicago, IL March 21   08:00 AM</p> <p>Now GenAI is for everyone. Discover the platform that makes it possible.</p> <p><a href="#">Event Details →</a></p>
 <p>Virtual Event</p> <p><b>LLM Fundamentals – Developing Your First LLM App</b></p> <p>Virtual Event March 22   12:00 PM</p> <p>Build, iterate and polish LLM ideas into production-ready solutions.</p> <p><a href="#">Register now →</a></p>	 <p>Conference</p> <p><b>Microsoft Fabric Community Conference</b></p> <p>Las Vegas, NV March 26   08:00 AM</p> <p>Join Databricks at the Microsoft Fabric Community Conference March 26–28 in...</p> <p><a href="#">Register Now →</a></p>	 <p>Regional Event</p> <p><b>Data Intelligence Day Seattle</b></p> <p>Seattle, WA March 28   08:00 AM</p> <p>Now GenAI is for everyone. Discover the platform that makes it possible.</p> <p><a href="#">Event Details →</a></p>