Brought to you by:

databricks

The Data Intelligence Platform

for dummies®

A Wiley Brand

Democratize data & AI with intelligence

Understand enterprise data with AI

Innovate with ETL, DW, OLTP, BI, & AI

2nd Databricks Special Edition

Ari Kaplan
Amit Kara

About Databricks

Databricks is the Data and AI company. Thousands of organizations worldwide — including Comcast, Condé Nast, Grammarly, and over 60 percent of the Fortune 500 — rely on the Databricks Data Intelligence Platform to unify and democratize data, analytics, and AI. Databricks is headquartered in San Francisco, with offices around the globe, and was founded by the original creators of Lakehouse, Apache Spark™, Delta Lake, and MLflow. To learn more, follow Databricks on social media:

 **x.com/databricks**

 **linkedin.com/company/databricks**

 **facebook.com/databricksinc**

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

The Data Intelligence Platform

2nd Databricks Special Edition

by Ari Kaplan and Amit Kara

for dummies®
A Wiley Brand

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

The Data Intelligence Platform For Dummies®, 2nd Databricks Special Edition

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

Copyright © 2026 by John Wiley & Sons, Inc., Hoboken, New Jersey. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permissions.

Trademarks: Wiley, For Dummies, the Dummies Man logo, The Dummies Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. Databricks and the Databricks logo are registered trademarks of Databricks. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services, or how to create a custom *For Dummies* book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact info@dummies.biz, or visit www.wiley.com/go/custompub. For information about licensing the *For Dummies* brand for products or services, contact BrandedRights&Licenses@Wiley.com.

ISBN: 978-1-394-39641-2 (pbk); ISBN: 978-1-394-39642-9 (ebk); ISBN: 978-1-394-39643-6 (ePub). Some blank pages in the print version may not be included in the ePDF version.

Publisher's Acknowledgments

Some of the people who helped bring this book to market include the following:

Developmental Editor and Project Manager: Carrie Burchfield-Leighton

Contributing Writer: Stephanie Diamond

Sr. Managing Editor: Rev Mengle

Acquisitions Editor: Traci Martin

Sales Manager: Molly Daugherty

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Table of Contents

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Introduction

Your organization's success relies on the effective use of data to drive intelligent decision-making and business impact. Data intelligence has fundamentally improved what's possible with your data, enabling your organization to build applications that reason on your own data, with humans in the loop. Data intelligence gives you the power to make smarter decisions, operate more efficiently, and achieve greater success.

The Databricks Data Intelligence Platform unlocks entirely new data and AI use cases, powering data-intelligent applications with open data formats, unified governance, and composable agents that know your business. It's built on a data lakehouse, which provides a unified platform for data and AI stacks, which helps your organization democratize data, build AI applications, better align with shared metrics, break down data silos, and establish operating models to move from pilots into the core of the business. Fully capitalize on your own data assets, leveraging traditional and generative AI (GenAI), data warehousing, business intelligence, online transaction processing, and governance.

About This Book

The Data Intelligence Platform For Dummies, 2nd Databricks Special Edition, explores how you can shift from reactive to proactive strategies and use your data as a competitive asset. This book covers

- The value of data intelligence and the power of AI
- The Databricks Data Intelligence Platform
- Using traditional and GenAI to build applications

Foolish Assumptions

In writing this book, we made a few assumptions about you:

- You want to leverage AI to solve complex problems, and you want solutions that integrate AI with data.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

- You seek a unified, open, and scalable platform to drive efficiency, innovation, and a competitive advantage.
- You're responsible for ensuring data governance, security, and regulatory compliance.
- You want to learn more about the Databricks platform and how it integrates with your existing infrastructures.

If any of these assumptions describe you, you've come to the right place.

Icons Used in This Book

Throughout this book, different icons are used to highlight important information. Here's what they mean:


TIP

The Tip icon highlights information that can make doing things easier or faster.


REMEMBER

The Remember icon points out things you need to remember when searching your memory bank.


WARNING

The Warning icon alerts you to things that can harm you or your company.

Beyond the Book

This book can help you discover more about data intelligence platforms, but if you want resources beyond what this book offers, here are more insights:

- See demos, product tours, videos, and tutorials on Databricks: `databricks.com/resources/demos`.
- Join your peers in the 100,000+ strong Databricks community: `community.databricks.com`.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

IN THIS CHAPTER

- Exploring the value of data intelligence
- Learning about the key features of data intelligence platforms
- Finding out about use cases in several industries

Chapter 1
Understanding Data Intelligence

When utilized effectively, data intelligence has the potential to revolutionize decision-making and democratize data interaction for everyone. It allows nontechnical users to converse with their own data in natural language in the context of their business. Data intelligence takes business to the next level by leveraging generative artificial intelligence (GenAI) with modern data warehousing, online transaction processing (OLTP), business intelligence, and data science to get more intelligent insights and deliver strategic decision-making.

REMEMBER

GenAI is any type of AI capable of interpreting data or creating new content by itself. GenAI content encompasses a wide range of formats (structured, unstructured, semi-structured) and workflows (data warehousing, OLTP, databases, batch, real-time). AI can analyze information, make decisions, and take actions to achieve specific goals, freeing up time and resources for strategic initiatives.

Data intelligence refers to the application of AI to understand the context of your organization's data, to self-serve much more actionable insights, and to enable your employees to work more efficiently and intelligently. This process leverages GenAI to sift

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

through and make sense of vast amounts of your data, helping you derive intelligent insights that can inform decision-making and improve services, investments, and overall business strategies.

This chapter explores the definition, benefits, and impact of data intelligence.

Learning about Data Intelligence

Companies that want to gain a competitive advantage need to make sense of their data. Augmented with GenAI, data intelligence encompasses a range of activities, from data gathering and analysis to applying data insights to solve real-world problems. By leveraging data intelligence, companies can more easily uncover complex patterns, predict trends, and make evidence-based decisions. This section examines how data intelligence gives companies the tools they need to succeed.

Intelligent

Data intelligence combines GenAI with the unification benefits of a lakehouse architecture to power and govern data intelligence that understands the unique semantics of your data. This data foundation enables data intelligent applications that can reason over your enterprise data, tailored to your business.

REMEMBER

The Databricks Data Intelligence Platform unlocks entirely new data and AI use cases, which power data-intelligent applications with open data formats, unified governance, and composable agents that know your business. This platform also automatically optimizes performance and manages infrastructure, tailored to your business. For more information on the Databricks Data Intelligence Platform, see Chapter 3.

Simple

Natural language simplifies the user experience. Data intelligence understands your organization's language, so search and discovery of new data is as easy as asking a question like you would to a coworker. Additionally, developing new data and applications is accelerated through natural language assistance, which enables writing code, remedying errors, and finding answers.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Governed

Data and AI applications require robust governance and security, particularly with the emergence of GenAI. Databricks provides these end-to-end capabilities through Unity Catalog. You get transparency every step of the way — from the raw data federated across your entire data estate, through everything downstream such as applications, notebooks, dashboards, and reports. You can securely share or protect data with federation, marketplaces, and cleanrooms without compromising data privacy and IP control. This capability is particularly important for enterprises, such as medical or financial organizations, that handle sensitive information.

TIP

Without data intelligence, platforms aren't smart. Platforms like legacy data warehouses (DWs) aren't intelligent because they simply store data and lack semantic understanding of the data, which requires highly skilled engineers to manually maintain infrastructure and experts to write code to retrieve information. Modern DWs continually learn about the usage trends and human feedback and then apply the learnings to make the entire platform better and more efficient.

Robust

Companies seeking a competitive edge must effectively gather and interpret their data. Data intelligence helps better collect and combine data into robust datasets, analyzes the data with analytics and AI, and helps with real-world decisions.

Streamlined

Data intelligence is key for your company to make sense of its data. Using advanced tools helps you better understand your customers and the market and make smarter decisions. Here are a few examples of how data intelligence streamlines this:

- **Enhancing data democratization:** Through the use of natural language, decision-makers can now independently ask questions about their data without the help of more technical programmers. This enables significantly more people to get value from a company's data assets.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

- **Streamlining operations:** Data intelligence automatically optimizes performance and manages infrastructure in ways unique to your business.
- **Ensuring data governance and compliance:** Understanding your data also means knowing where it comes from, figuring out how you'll use it, and making sure it complies with legal and ethical standards. Data intelligence provides the tools for effective data governance, enabling companies to manage data quality and security, ensuring compliance with regulations.

Unified

Unified, open, and scalable lakehouse architectures built with data intelligence serve as a comprehensive system that integrates data-related functions into a single, cohesive environment.


TIP

Organizations benefit from a more efficient data management and analysis approach by leveraging unified platforms. These platforms eliminate data silos and provide a single, centralized repository for all data assets. This ensures consistency, accuracy, and governance across the organization.

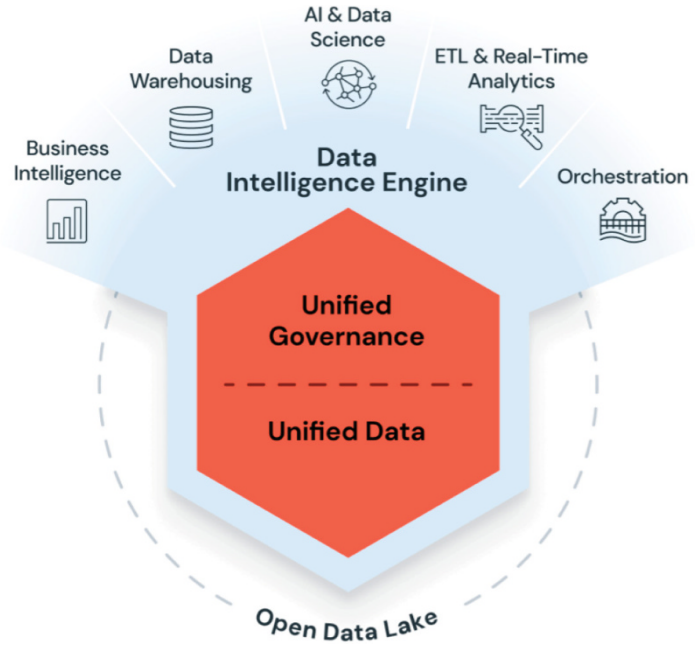
Maximizing Benefits from Data Intelligence

Data intelligence has developed as a key strategy for organizations, helping them utilize the power of their data. Data intelligence simplifies development that doesn't require skilled personnel. This section outlines some benefits these initiatives can provide your organization.

Making data easily searchable and understandable

Data intelligence understands your organization's language, so searching and discovering relevant data is as easy as asking a question like you would to a coworker. As shown in Figure 1-1, data intelligence makes information easily discoverable and searchable by understanding the context of the question — and not just a simple keyword match. Additionally, natural language processing (NLP) allows users to query data using plain language.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.



Source: Databricks

FIGURE 1-1: Data intelligence engines add unified governance and data in every step of the end-to-end data platform.

REMEMBER

Natural language is a critical feature of data intelligence and enables people and AI agents to simply talk with and reason on your vast data estate. A centralized semantic layer can define every table, every column of data, business definitions, and business information, making all interactions much more relevant and helpful to the entire decision-making process.

Unifying siloed data into a single platform

Unifying siloed data into a single platform addresses an issue many organizations face — data fragmentation across different systems, departments, and locations. When data is siloed, it's isolated from other relevant data, making it almost impossible to understand concepts such as customer behavior or market trends.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

WARNING

Fragmented data can lead to inefficiencies and, more importantly, missed opportunities because managers can't see the big picture. Combining data into a single unified platform helps companies break down these silos, allowing data to flow and be analyzed.

REMEMBER

A unified data platform provides a centralized repository where all data (structured or unstructured) can be stored and analyzed. This ensures that data is accurate and allows advanced analytics and AI applications to be used more effectively. As a result, organizations can leverage their data assets to the fullest.

Empowering non-technical users to get data insights

Providing simpler access to data means making it more accessible for nontechnical users. This enables them to get insights without relying on IT departments. Make sure that all employees understand the basics of data analysis and the tools available that make it so simple.

Streamlining company operations and cost savings

Data intelligence can streamline company technology operations that lead to cost savings. Predictive analytics forecast trends and allow companies to adjust their strategies.

TIP

AI can uncover new opportunities for technology efficiencies and cost reduction, such as automating time-consuming manual processes. Examples include identifying and reallocating improperly utilized resources, or reorganizing how data is stored so it can scale larger and faster. The key is that AI can now improve the operations of every technology step from raw data software engineering to analytics.

Fostering collaboration

Data intelligence tools facilitate collaboration across different teams by providing a common environment into all tasks. Teams can work simultaneously on the same datasets, develop code together, share dashboard and reporting insights, and make collective decisions. This collaborative environment encourages people to work toward common goals.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Impacting the Entire Business

Data intelligence improves the functionality and efficiency of every aspect of business. It drives their evolution, ensuring that they respond more to human needs and ethical standards. This enhances the entire data and AI ecosystem, making it capable of addressing complex challenges. This section discusses key ways data intelligence shapes the landscape.

Improving data quality

The quality of data is foundational to the effectiveness of AI applications and analytical processes. Data intelligence enhances these aspects by providing mechanisms for managing the quality of data and AI agents and consistent management across different data sources.

Driving innovation and new business models

Data intelligence is crucial in delivering innovation and shaping new business models. Companies can identify emerging trends and underserved market needs by analyzing their data, which opens up opportunities for new and innovative products and services.


TIP

A data-driven approach allows businesses to experiment with business models, such as subscription services or on-demand platforms, which can provide a competitive edge. Insights gained from data intelligence platforms can lead to new revenue streams and transformative strategies.

Accelerating AI and ML

Data intelligence provides the foundation for AI and machine learning (ML) by preparing and transforming data into a format these technologies can use. High-quality, well-governed data is essential for training accurate and reliable AI and traditional ML models.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Evaluating Key Features of Data Intelligence Platforms

Data intelligence platforms give businesses the tools to unleash their data into valuable business assets like never before. They're founded on a unified lakehouse to analyze data and formulate effective strategies. Knowing what these platforms can do helps you choose the right one for your data needs and goals.


TIP

When evaluating data intelligence platforms, consider factors such as cost, scalability, performance, ease of use, and open-source and integration capabilities to ensure that the chosen platform aligns with your immediate business requirements. The platform should also be futureproof, with the agility to incorporate the dizzying levels of new industry AI capabilities and to innovate along with the market.

Making the platform usable for diverse skill levels

Data intelligence platforms should be accessible to people with varying levels of technical expertise to ensure that a range of users — from data scientists to business analysts to executives — can leverage the platform's capabilities.


TIP

Simplifying the experience of non-technical users so they can self-serve insights and dashboards without waiting on technical staff empowers stakeholders to make more informed data-driven decisions. Also, technical users should have a simpler and more powerful experience to develop code, workflows, and applications.

Automating data processes

Automation in data intelligence platforms transforms how companies handle their vast amounts of data. Businesses can significantly enhance efficiency, accuracy, and speed by integrating automation into their data processes. Automation streamlines workflows, reduces manual intervention, and improves the overall data management experience. And with the right platform, you can get a huge lift in productivity out of the box.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

REMEMBER

One of the most important benefits of automation is that it reduces the need for manual data handling and workflow management. Manual tasks are time-consuming and prone to errors. Automation incorporates human feedback throughout the process, reducing risks while dramatically improving and accelerating business.

TIP

Businesses can improve operational efficiency, save time, and apply resources to strategic initiatives by automating tasks like data collection, cleaning, and processing.

Examining Data Intelligence Use Cases in Diverse Industries

Data intelligence is used across every industry from finance to healthcare to energy. Using data-driven insights and leveraging AI improves and can fundamentally change how businesses operate. While endless use cases across sectors exemplify how data intelligence helps companies learn about their customers, improve processes, and spot fraud, we provide a few examples below:

- **Financial Services:** This sector uses data intelligence to handle financial risks, predict economic trends, and follow regulations. Banks and other financial institutions use AI to assess creditworthiness, spot fraud, create customer-facing support agents, and provide smarter customer market intelligence.
- **Healthcare and Life Science:** Healthcare organizations use data intelligence to improve patient care, control costs, and accelerate clinical trial research. Data intelligence better informs medical decisions, identifies individualized drug targets, accelerates clinical trial research, and streamlines reimbursements from claims appeals.
- **Media and Entertainment:** These industries use AI-driven applications to contextualize ad targeting, auto-generate localized content, build autonomous network operations, and prevent fraud from anomaly detection.
- **Retail and Consumer Goods:** They apply data intelligence to hyper-personalize offers at scale, optimize inventory more granularly, streamline supply chain logistics, and understand commerce through agentic searches.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

- **Manufacturing and Auto:** These industries use AI to get better understand what maintenance actions are needed, automate much of the order processing, and improve field services.
- **Insurance:** Insurance companies use data intelligence to evaluate risks, set prices for insurance plans, and find false claims. By studying large amounts of data, they can get a clearer picture of risks and make filing claims more efficient.
- **Energy and Utilities:** Energy companies use AI for predictive power outage prevention, asset reliability to reduce disruption, hyper-personalized energy savings recommendations, and regulatory compliance co-pilots.

REMEMBER

Data intelligence applications may vary across industries, but the common goal remains the same. It's to extract valuable insights from data and leverage them to drive business growth and enhance customer experiences.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

IN THIS CHAPTER

- Looking at the challenges of not having a lakehouse
- Exploring lakehouses, legacy data warehouses, and data lakes
- Looking at GenAI and classical AI
- Understanding the potential of AI
- Using lakehouse architecture for GenAI
- Deploying data intelligence platforms

Chapter 2
Exploring the Lakehouse as the Foundation for Data and AI

Data intelligence platforms have a data lakehouse architecture as their foundation, augmented with generative artificial intelligence (GenAI), that offers powerful ways to democratize data and AI across your organization. Lakehouse architectures store, process, and govern huge amounts of structured and unstructured data together into one unified environment, bringing data warehousing, online transaction processing (OLTP), business intelligence (BI), classical AI, and GenAI to new heights.

This chapter examines the value of a lakehouse architecture and how incorporating GenAI and classical AI into a lakehouse greatly enhances its value for your organization.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Experiencing Challenges without a Lakehouse

Success always begins with data, but without a lakehouse, the data and AI estate is fragmented. Most companies struggle to effectively integrate and govern data and AI to achieve their business objectives. These challenges incorporate various components, each critical to the data intelligence ecosystem:

- Your data and AI are siloed. Data silos drive high operational costs.
- Your data privacy and controls are challenged. Inconsistent policies reduce the trust in the data.
- You depend on highly technical staff. Disparate tools slow down cross-team production.

This mess of disparate and complex systems has to be stitched together, making it hard to succeed. Each component has siloed metadata with differing definitions and its own access controls, license fees, and security — all lead to slower decision-making for organizations, higher costs, and constantly copying data back and forth. Here are some of those components and their challenges (also shown in Figure 2-1):

- **Data warehouses (DWs)** store large volumes of historical data in a structured format for business intelligence. Challenges include high costs, complexity, and the inability to handle unstructured data.
- **Data lakes** are centralized repositories of raw data from diverse sources, such as unstructured and semi-structured data. The challenge lies in data reliability with managing and governing huge, unstructured datasets.
- **GenAI** helps build applications that reason, create, and interact with the nuance of a human expert. Challenges abound with having GenAI gain intelligence from your specific data, govern it, and maintain it for accuracy and reliability.
- **Transactional databases** process many small, concurrent transactions in real time. Challenges include being locked in to expensive, on-premises legacy databases, which aren't built for GenAI.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

- **Orchestration and extract, transform, load (ETL)** brings in raw data and transforms it through business rules to a more usable format. Challenges include handling the complexities of large, complex, and diverse data sources and the daunting amount of manual work required to coordinate it all.
- **Streaming** data is real-time information, such as social media or sensor data, that's processed continuously. Challenges are that traditional DWs aren't able to handle streaming data, and that managing the quality and governance of data as it is ingested is difficult at scale.
- **Machine learning (ML)** is classical AI, making predictions and classifications with structured data. Challenges include needing high-quality data as the foundation and specialized talent to build, deploy, and monitor ML models.
- **Business intelligence (BI)** provides insights, typically in reports and dashboard on historical, structured data such as sales history. Challenges are that business users can't easily self-serve their own insights and instead rely on technical staff to make it happen. Also, there's extra software license costs and difficulty to discover the relevant data across thousands of sources.
- **Data science** extracts valuable insights from datasets to uncover patterns from vast information and the complexities of information. Challenges are properly governing the added complexity, getting a unified data foundation of structured and unstructured data, a lack of specialized talent, and governance.

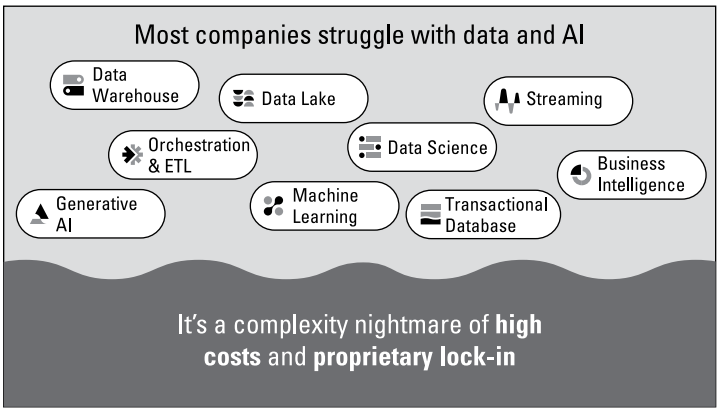


FIGURE 2-1: The challenges of the data intelligence ecosystem.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

To add to the challenges themselves, you have to navigate the unified governance challenges to navigate regulations and implement strong data management and security controls. For more details about these components and how the Databricks Data Intelligence Platform solves these challenges, see Chapter 3.

REMEMBER

A data intelligence platform has many elements that help an entire organization: an open data lake for storing and managing all your data types; unified data storage for reliability and sharing; a unified security, governance, and catalog environment; and an AI-powered engine to understand the semantics of your data. A data intelligence platform improves the experiences of data science and AI, ETL, and real-time analytics, orchestration, and data warehousing.

Comparing Lakehouses with Legacy Data Warehouses and Data Lakes

Lakehouses represent a distinct approach to data storage and analytics from DWs and data lakes. The transition from legacy DWs and data lakes to lakehouses was prompted by the need for more scalable, open, and cost-effective solutions for managing huge amounts of structured and unstructured data. In this section, you look at the differences.

Open architectures

With Databricks, your data is always under your control, free from proprietary formats and closed ecosystems. The Databricks lakehouse is underpinned by widely adopted open-source projects Apache Spark, Apache Iceberg, Delta Lake, MLflow, and Unity Catalog. On top of this, Delta Sharing offers an open solution to securely share live data from your lakehouse to any computing platform without costly replication and complicated ETL.

Unified architecture

Lakehouse architecture unifies all integration, storage, processing, governance, sharing, analytics, and AI. It's the leading approach to working with structured and unstructured data; it's one end-to-end view of data lineage and provenance that gives

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

you the ability to use any language (SQL, Python, R, Scala) and to batch or stream data to the lakehouse. It's one platform for all three major cloud providers.

Scalable

Lakehouses are more scalable than legacy DWs and data lakes; they scale up to trillions of records with lower cost and higher performance. They offer automatic optimization for performance, and storage ensures the lowest total cost of ownership (TCO) of any data platform together with world-record-setting performance.

Improving data governance and security

Lakehouses improve how data is governed and protected by using one security and governance model for all data and AI access across the organization. Having one unified governance platform makes it easier to follow regulations and keep data safer than legacy DWs and data lakes because they have multiple disjointed governance solutions and different kinds of data, making it more difficult to apply consistent policies and protections.

Distinguishing between GenAI and Classical AI

Gen AI and classical AI are two branches of AI. Classical AI is useful in fields where you need to make numerical predictions such as predicting future sales across each store, or to classify items such as grouping your millions of customers into different segments. On the other hand, GenAI offers value in generating new content such as text summaries, images, code, and synthetic data.

GenAI generates new content based on the patterns it learned from its training data. Instead of simply analyzing data, GenAI systems can interpret and search through text, images, audio, video, and other media. GenAI can also be used internally to create new text and to write and edit software code.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

REMEMBER

A key capability of GenAI is agents that can autonomously perform complex tasks, enhance decision-making, and improve operational efficiency. Some examples of GenAI for business include

- **Assisting with software development:** Can take prompts in natural language and write code in languages such as SQL, Python, Scala, and R
- **Documenting data assets:** Can describe the contents of a table and columns for better data discovery through semantic searches that understand the context of your questions
- **Knowledge assistance:** Can turn your company's own data into expert AI chatbots
- **Extracting information:** Can take large amounts of existing documents and give a synopsis or a grade for easier interpretations from humans

Realizing the Significance of AI in Enhancing Data Intelligence

Combining AI with data intelligence creates great advancements in how businesses analyze, understand, and leverage their data. This impact boosts the strategic capabilities of organizations. It allows companies to adapt more quickly to market changes and consumer needs.

REMEMBER

The importance of AI in improving data intelligence lies not just in its technological capabilities but in its ability to drive innovation across various sectors.

Employing the Use of Lakehouse Architecture for GenAI

Integrating lakehouse architecture with GenAI enhances the capabilities of both technologies. This integration creates a more powerful environment for data processing and analytics.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Leveraging a lakehouse architecture

Lakehouse architecture provides a data storage and management foundation by combining the best features of data lakes and data warehouses. It enables organizations to store structured and unstructured data in a single repository while still being able to perform analytics and ML tasks.

REMEMBER

Incorporating GenAI introduces new ways to analyze and generate data-driven insights. This technology can help organizations improve data quality and develop more accurate predictive models, which is a competitive advantage.

By integrating lakehouse architecture and GenAI, organizations can manage their data more effectively and find new possibilities for data-driven decision-making.

Utilizing open data storage

Utilizing an open data storage helps with reliability and data sharing. It's essential for employing lakehouse architecture and incorporating GenAI capabilities. It's a technology that provides an efficient data storage layer, helping organizations develop and deploy GenAI applications easily and efficiently. By leveraging this, organizations can use the power of GenAI to best drive innovation.

Integrating GenAI capabilities into a lakehouse

Integrating GenAI capabilities into a lakehouse architecture enhances data analysis. By leveraging the combined strengths of a lakehouse, organizations can elevate their data intelligence efforts. A few key benefits of this integration are

- **Automating data tasks:** GenAI can streamline data operations within a lakehouse. While classical AI may automate data cleansing, GenAI can further assist by generating artificial data for testing and training models, ensuring robust analytics and AI applications.
- **Enhancing data discovery:** AI enhances intelligent search capabilities within a lakehouse. Users can utilize natural language queries to efficiently discover and comprehend

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

the relationships between data assets. This simplifies data discovery beyond a mere keyword search and ensures that the right datasets are easily accessible for analysis.

- **Developing custom AI applications:** Integrating AI into the lakehouse framework allows organizations to create applications tailored to specific needs. Examples include making large language models (LLMs) on your own company's data, developing predictive models, customizing recommendation engines, or automating complex reporting tasks.

REMEMBER

Integrating GenAI capabilities into a lakehouse architecture empowers organizations to unlock entirely new use cases and extract more value from their data.

Enabling data teams to collaborate

The combination of lakehouse architecture and GenAI capabilities significantly improves the collaborative potential of data teams. It enables a more dynamic exchange of ideas. This fosters a culture of innovation. Data teams can work together to build, train, and deploy AI models more efficiently. This leverages the strengths of a lakehouse's data management and the creative potential of GenAI to drive business growth.

Enhancing data analysis and insights with AI

To enhance data analysis with AI, leverage various AI-powered tools and techniques to automate and streamline multiple stages of the data analysis process. AI can be integrated into different phases of data analysis in the following ways:

- **Data preparation:** AI can automate the data preparation phase, which includes cleaning, organizing, and preprocessing data. AI tools can detect and correct data quality issues, extract information from unstructured data, and combine data from different formats.
- **Data exploration:** AI algorithms can explore your data by using natural language. This helps uncover insights that may not be apparent to humans.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

- **Data interpretation:** AI can improve data interpretation by generating summaries, insights, or stories from your data. It can identify causal relationships and predict future outcomes or actions based on the data.
- **Data quality:** AI can help detect when data and model quality is skewed, automatically flag it, and assist in remediation.

Automating complex data tasks and processes

Automating complex data-related tasks and processes means making the handling and analysis of data more efficient by using technology to do the work. This organizes large amounts of unstructured data in data lakes, builds and applies GenAI and ML models, and handles continuous data flow in real time.

For orchestrating jobs, AI can automatically select the right instances and start time to hit your requirements. It handles tasks like auto-scaling and error remediation for you.

Many aspects of data engineering, such as optimizing file sizes for tables, can benefit from AI and be automated as well. Engineers typically invest a lot of their time and expertise to figure out the optimal file sizes for reading or writing data, which can lead to substantial performance improvements. Automating this complex task is a game-changer.

Intelligent autoscaling in ETL processing optimizes cluster utilization and minimizes end-to-end latency for streaming workloads by automatically adjusting resources based on data volumes and processing needs, up to a specified limit. It efficiently scales up when data arrival outpaces processing and scales down during low load, ensuring task completion before shutting down to save on infrastructure costs.

Deploying a Data Intelligence Platform

A data intelligence platform helps organizations innovate with a unified platform that combines uses across various personas, including data scientists, data engineers, architects, and business

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

analysts. It combines different stages of data into a single, integrated environment. The following enables this integration:

- **Integrating data:** You can bring data from different sources into a single place. It supports data integration from transactional databases, data warehouses, data lakes, and streaming data sources, making it easier to work with all your data on a single platform.
- **Processing and analysis:** After your data is in the platform, you can process and analyze it. The platform supports all major languages you may prefer, such as Python, R, Scala, and SQL. Using built-in functions and libraries, you can easily clean, transform, and analyze your data.
- **Collaborating in a workspace:** The platform gives you shared workspaces where different team members can work together on the same data and projects. Data engineers, data scientists, business analysts, and non-technical workers can create visualizations and dashboards, all on the same platform. Shared workspaces help maintain version control, so everyone uses the most current data and analyses.
- **Using a unified governance:** You can manage your entire data analytics workflow from a single place to control access to data, manage resources, and monitor jobs in one place. This allows better resource allocation and monitoring of ongoing jobs.
- **Deploying seamlessly:** After you've built your data analytics solution, you can easily deploy it to production. Seamless deployment enables organizations to move data projects from development to production without the typical issues that arise without having a data intelligence platform.

Check out Chapter 3 for more information on the Databricks Data Intelligence Platform.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

IN THIS CHAPTER

- Reviewing the Databricks Data Intelligence Platform
- Looking at the Databricks architecture
- Assisting developer efforts

Chapter 3
Getting Started with the Databricks Data Intelligence Platform

Companies are continually seeking ways to simplify their data architecture while enhancing their ability to derive meaningful insights. This chapter looks at the foundational aspects of getting started with the Databricks Data Intelligence Platform.

Introducing the Databricks Data Intelligence Platform

The Databricks Data Intelligence Platform allows your entire organization to use data and artificial intelligence (AI). It's built on a data lakehouse to provide an open, unified foundation for all your data, AI, and governance needs and is empowered by an understanding of your company's unique data.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Databricks helps you simplify and accelerate your data and AI goals and augments all your workloads — from extract, transform, load (ETL) to online transaction processing (OLTP) to data warehousing to generative AI (GenAI) — with more intelligence.

Enterprises everywhere are reimagining how they build intelligent applications, and they need a platform that can help them leverage their data to power AI applications and deploy quality agents that actually work.

Delivering data intelligence with Databricks

Databricks combines the power of GenAI with the comprehensive features of a lakehouse architecture to create an end-to-end platform. Databricks continually learns the unique nuances of your ever-changing business and data, powering a wide range of use cases. Any employee or end-user can search, understand, and query data by using natural language. Databricks uses information about your data, usage patterns, and trends to understand your business's jargon and your unique data environment. With purpose-built agents built on your specific data, you get significantly better insights than with the use of generalized large language models (LLMs) trained on a vast amount of data across the Internet.

Companies are adding AI assistants, but in reality, many of these solutions fall short because they aren't trained on their own data. Even simple definitions such as "Did a customer churn?" or "What day does our fiscal year start?" vary across companies and even across business units.

Databricks directly solves this problem by automatically learning about business and data concepts throughout your enterprise. Its intelligence comes from signals across your data estate, including centralized business semantics, table and column descriptions, dashboards, notebooks, data pipelines, usage, and human feedback. Because of Databricks' unique unified end-to-end nature, it can see how data is used in practice across the entire estate. This lets Databricks build highly accurate, specialized models and AI agents for your enterprise.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Ensuring privacy and governance

The need for strong governance and security in data and AI applications has never been more important. Databricks includes a comprehensive platform for machine learning operations (MLOps) and LLM operations (LLMOps) and AI development supported by a unified approach to governance and security. The platform allows you to pursue a wide range of AI initiatives, enabling you to maintain privacy and control over your intellectual property.

REMEMBER

MLOps is a function of ML engineering. It's focused on streamlining the process of taking ML models to production and maintaining and monitoring them as your data changes.

Using the Databricks Data Intelligence Platform

Databricks has created a data intelligence platform that utilizes the power of the data lakehouse and GenAI. The Databricks Data Intelligence Platform stands out because its unified governance layer covers both data and AI. It also has a single query engine that spans ETL, SQL, ML, and business intelligence (BI). This integration is crucial in making data accessible to everyone in your organization.

REMEMBER

Databricks pioneered the lakehouse concept. It provides an open, unified architecture for all data — structured and unstructured — and its governance needs.

TIP

Databricks developed performance enhancements like Photon (the next-generation engine that provides extremely fast query performance at low cost), which makes the platform more scalable and efficient. This ensures that Databricks can handle even the largest-scale data workloads.

A visual representation of the architecture of the Databricks Data Intelligence Platform is shown in Figure 3-1. Starting at the bottom of the figure, this section looks at each component to see how everything fits together.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

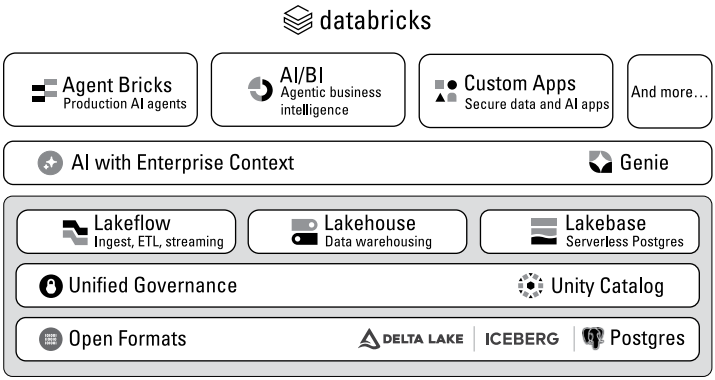


FIGURE 3-1: The Databricks Data Intelligence Platform.

Open Data Lake

With Databricks, your data is always under your control, free from proprietary formats and closed ecosystems. The data lake can be used to store, refine, analyze, and access a wide variety of data formats — whether structured, semi-structured, unstructured, batch, or real-time streaming. The Databricks lakehouse lets you choose your favorite open source data storage format, including the two most popular on the market: Delta Lake and Apache Iceberg. There is no need to copy data over and over again among these formats, and no need to lock-in on any single format choice either.

Furthermore, Databricks uses AI to optimize data storage challenges, so you can get faster performance without having to always manually manage tables and volumes, even as your data changes over time.

REMEMBER

Data storage has two major open formats: Delta Lake and Apache Iceberg. In the past, companies duplicated and had multiple copies of their data in several places and formats. This practice is costly and time-consuming and doubles your cost and effort.

Unity Catalog

Databricks Unity Catalog (UC) offers a unified governance layer for data and AI within the Databricks Data Intelligence Platform. With UC, organizations can seamlessly govern their structured

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

and unstructured data, ML models, notebooks, dashboards, and files on any cloud or platform. Data scientists, analysts, and engineers can use UC to securely discover, access, and collaborate on trusted data and AI assets, leveraging AI to boost productivity and unlock the full potential of the lakehouse architecture. This unified approach to governance accelerates data and AI initiatives while simplifying regulatory compliance.

UC has a centralized catalog of your business semantics and metrics. It can automatically define descriptions and tags of all data assets in UC with optional guidance from humans. These assets are then leveraged to make the whole platform aware of jargon, acronyms, business rules, metrics, and semantics. This process enables better semantic search, better AI assistant quality, and an improved way to govern data and applications.

Databricks also significantly elevates searching for data from a simple keyword search to a fully contextual search. Databricks doesn't just find data; it interprets, aligns, and presents it in an actionable, contextual format, helping all users get started faster with their data, or coders find the data they need among thousands of similar sounding table and column names. For example, Databricks can help users determine which table has data from last week or last year, and if "sales_NE" means sales in Nebraska, New England, or the Netherlands.


TIP

After assets are registered in UC, Databricks significantly improves the discoverability of data by allowing users to search for data using natural language and company-specific terminology, making it easier for users to use data assets within the organization.

Agent Bricks

Databricks gives you Agent Bricks, a robust environment to build, productionize, monitor, and evaluate all your composable agent needs. It comes out of the box with a variety of pre-built agents to get you started, and the ability to easily create and manage your own custom agents, often by simply starting with the business problem you are trying to solve. Importantly, Agent Bricks does all the orchestration behind the scenes, pulling data, doing the judging, all while maintaining governance, cost, and helping you with balancing cost vs quality.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

REMEMBER

For more about Agent Bricks and how the platform enables you to build your own GenAI applications, see Chapter 4.

Databricks Lakeflow and Spark Declarative Pipelines

Data engineering is at the core of everything in data and AI. For analytics and AI to be effective, companies must reliably ingest raw data and transform it into formats that downstream applications can utilize, such as reports and dashboards. This is where Databricks Lakeflow shines, helping data engineers create and manage repeatable end-to-end data pipelines and version-controlled workflows — whether raw data intake or automated. This capability allows you to run your jobs without pre-configuring and managing the underlying infrastructure. Simply define the transformations to perform on your data and let DLT pipelines automatically manages task orchestration, cluster management, monitoring, data quality, and error handling.

Databricks Lakeflow is the declarative ETL framework for the Databricks Data Intelligence platform for batch and streaming procuess with three key components: Lakeflow Connect for scalable data ingestion (ETL/ELT), Spark Declarative Pipelines (SDP) for simplified and reliable data transformation, and Lakeflow Jobs for orchestrating and scheduling these pipelines as workflows with all the appropriate monitoring and observability.

Databricks Workflows orchestrates data processing, ML, and analytics pipelines on the Databricks Data Intelligence Platform. With a wide range of task types, deep observability capabilities and high reliability provide your data teams with the tools to better automate and orchestrate any pipeline on serverless compute. It automatically optimizes and scales resources for your workloads and starts up environments nearly instantly, making implementing data processing and analysis pipelines easier. This all keeps costs low and performance high.

REMEMBER

ETL is the process data engineers use to extract data from different sources. Then, they transform the data into a usable and trusted resource. Finally, they load that data into the systems that end-users can access and use downstream to solve business problems.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Databricks SQL

Databricks SQL is one of the leading serverless data warehouses. A few of its capabilities include

- Running your ETL workloads and BI, with the added benefit of governance through UC
- Using open-source foundational architecture that scales with optimal price and performance
- Being built for AI and the modern data stack
- Using advanced techniques to speed up data access

Lakebase

Transactional databases store structured data, ensuring data reliability by managing individual records like sales records, bank transactions, and support call logs. However, legacy databases have been around for over 40 years and haven't really changed their technology since then. There is massive vendor lock-in from data being so sticky. They are hard to migrate. They are very expensive, with the need to pay for many copies of databases. And they are hard to use, taking lots of time to spin up, and full-time administrators to manage. They are built for being on-prem and, perhaps most importantly, weren't built for AI. You can't deliver next-generation applications with legacy databases.

Databricks' Lakebase is a new category of databases built for the agentic era. This category is defined by a few key innovations:

- Its open-source Postgres foundation helps you avoid lock-in, and full support for community extensions lets you access a massive ecosystem of innovation.
- There is a complete separation of compute and storage, which provides very low latency, high queries per second (milliseconds per query), and production SLAs for Lakebase.
- It's built for AI, operating at the speed of agents. Database instances can be launched under a second, and you truly pay only for the compute that you use.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

AI/BI

Companies want to use data intelligence to then extend BI to every team. After you have great data, you need to get value out of it. This is where AI/BI comes in to marry BI with the power of data intelligence.

AI/BI Dashboards let anyone self-serve their own visualizations based on your data, getting useful results with AI-powered understanding. BI analysts can use natural language to create the data sets and reports their dashboards need. Plus, they can develop new visuals with natural language prompts or point-and-click.

AI/BI Genie is a chatbot that lets you simply talk with your data to explore any question. It responds in natural language, data tables, and visualizations as appropriate. Making a Genie Space is easy and near-instant, and it adheres to all the underlying governance through UC, ensuring secure data access.

Genie translates human-provided questions into analytical SQL queries, offering self-service data exploration, insights, and summaries. Best of all, it continually learns from user feedback to improve over time. What used to take months of effort to create a chatbot on your own data now takes minutes, without code.

Databricks Apps

Democratizing agents happen through applications. Every platform becomes democratized when you build applications on it, and these days, vibecoding is accelerating this faster than ever. But the biggest challenge is how to get governance working for production data across your applications. Agentic applications are hard to productionize when security and governance aren't negotiable. This is the hard part of scaling out apps.

Databricks Apps is a capability that enables developers to build and deploy secure data and AI applications directly on the Databricks platform, which eliminates the need for separate infrastructure. Basically, you create and bring your apps to your data. It's powered by the Databricks Data Intelligence Platform, integrates with everything Databricks offers, and provides security and governance through UC, built on an open ecosystem.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Data collaboration

Data sharing and collaboration are integral to the Databricks Data Intelligence Platform. Key building blocks include the following:

- Databricks Marketplace is an open third-party marketplace for all your data, analytics, and AI.
- Databricks Clean Rooms allows businesses to easily collaborate in a secure environment with their customers and partners on any cloud in a privacy-centric way.
- Delta Sharing — the most widely adopted open protocol for secure data sharing — powers both Marketplace and Clean Rooms. Delta Sharing lets organizations exchange live data and AI assets across platforms and clouds.
- Lakehouse Federation lets you use multiple data sources without having to migrate all the data into a unified system. Through UC, Databricks brings the compute right to the source (such as Snowflake, MySQL, Oracle, Hive) with zero copying of that data to duplicate locations. Because UC also handles open sharing, you can then expose all your data to tools such as Salesforce, Palantir, and SAP.
- UC secures and governs all the options in this list for data collaboration.

Using Databricks Assistant to Assist Programmers

Databricks Assistant is a context-aware AI assistant available natively in Databricks notebooks, SQL editor, and file editor. Databricks Assistant enables you to query data through a conversational interface, increasing your productivity within Databricks. You can describe your task in English and let the Assistant generate SQL queries, document complex code, and debug code. The Assistant leverages UC metadata to understand your tables, columns, descriptions, and popular data assets across your organization, providing personalized responses tailored to you.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

IN THIS CHAPTER

- Using AI to build applications
- Managing AI development challenges
- Looking into model management
- Building applications with AI and Databricks Apps
- Seeing it all come together

Chapter 4
Building AI Applications on The Databricks Data Intelligence Platform

Companies want to develop their own custom artificial intelligence (AI) applications to better serve their customers and stay ahead of the competition. Almost every facet of business is being influenced by AI: Workers are becoming much more efficient in their jobs, the majority of software development is being created with AI code assistants, and AI is unlocking new use cases across every industry. However, the task can be elusive to get AI that works well and understands your company's own enterprise data, guided by your business.

This chapter explores how the Databricks Data Intelligence Platform facilitates the development and deployment of AI applications. It delves into the platform's capabilities that enable seamless model building and the role of AI agents by using chatbots and dashboards to self-serve insights and building AI applications. The Databricks Data Intelligence Platform offers a unified solution for making the process of building, deploying, and monitoring all your AI solutions much easier. Databricks helps your AI needs, spanning predictive models to the latest generative AI and large language models (LLMs).

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Addressing the Challenges of AI Development

The challenge with building AI applications is it's hard. How do you build high-quality agents that provide not just the results that help your personal productivity but actually provide you with the results that are needed for you to deploy these agents in the path of risk?

Maybe you're in critical financial workflows or customer facing situations:

- **Building and standardizing without knowing what agents are doing is hard.** How do you evaluate the agents on your specific questions? What metrics are there? Do humans and LLM judges guide the process? How does the data drift over time?
- **Measuring and improving AI quality is hard.** So many AI techniques exist these days; it's a zoo of options for optimization. With new public models seemingly popping up every week, how do you know which one is right or built the right way? What is the strategy? Can you leverage research plans on how AI will approach resolving your question or job?
- **Governing and scaling safely is hard.** Finding the balance of cost versus quality is done through painful trial and error. How does the cost of AI scale over time, and will it be prohibitive as you grow?

The answer to all these challenges is to use a suite of features designed to streamline the AI development life cycle.

Considering Model Management and MLOps/LLMOps

AI models don't just pop into existence. An end-to-end process exists for building, deploying, and managing models, and the Databricks Data Intelligence Platform helps every step of the way. This section delves into the world of machine learning operations (MLOps) and LLM operations (LLMOps).

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Refining models

Models are built on data in the lakehouse environment — either directly in the Databricks Data Intelligence Platform or through connections to third-party environments. The Databricks Data Intelligence Platform ensures that data is quality, which leads to quality models. After your data has been transformed and loaded into the lakehouse, you can begin to make models.

Some models, such as linear regression, are better for predicting the optimal price of a product, for example, while other models, such as logistic regression, are better at predicting if a person should get a loan at a given rate or not.

Clarifying model explainability and transparency

People in the real world shouldn't blindly implement a recommendation from a model without understanding what drove the model's decision. A business demands to have visibility into which features are most important to drive the outcomes. For example, what were the reasons for and against the recommendation that a given prospect should be given a bank loan at a specific interest rate?

Without this explainability, the decisions are like black boxes, and human trust in the model is understandably limited. This often prevents models from ever seeing real-life practical use.

REMEMBER

Databricks offers full transparency from the raw data to the end results of the model with lineage tracking. This sheds transparency on every step of the data journey.

Deploying models

After you have the model, a data scientist or machine learning (ML) engineer can easily use Databricks to deploy it into production. This is achieved by creating a model serving endpoint — something that doesn't require much experience to accomplish. ML workflows help you move the model from experiment and development to staging and ultimately to production for use in the real world.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Observing model governance

Over time, a business grows from its first model in production to its second, to its 50th, and then before long, hundreds or thousands are in production. Being able to manage the life cycle of models is key, and Databricks makes this task simple.

Also, governing model life cycles through UC is critical. You need to make sure that only the right people can place models into production and that only the allowed users are able to access the data on which these models are built. Unity Catalog's audit logs and system tables show all these details, including who ran which model with what data and when.

Monitoring models and data drift

Your data is likely to change over time — interest rates fluctuate, shopping patterns change, and your customers' spending varies. This is known as *data drift,* and if data changes enough, such as product prices or new competitors entering the market, your models can be rendered less accurate or even worthless.

With the Databricks dashboards and underlying system tables, you can see the full health of each model, whether models are failing (for example, a data source stopped working), and whether a model needs to be refreshed. Metrics can be set during development with the option for custom metrics with lakehouse monitoring.

Developing AI Applications

AI is revolutionizing the way businesses in every industry develop new applications. This technology enables companies to accelerate innovation, tailor products, and address complex challenges. By adopting generative AI, businesses can shorten development times and enhance the scalability of their solutions. This makes it possible to deliver better services and products.

REMEMBER

Databricks' ecosystem lets you build AI applications from the ground up. You can start with raw data and develop AI models specifically designed to address the context of your business, and on your proprietary data, without leaking confidential

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

information outside your data perimeter. Some key AI features of the Databricks Data Intelligence Platform that enhance building AI applications include

- **Customizing agent training:** Databricks enables agents to be customized by using an organization's proprietary data. This ensures that the knowledge of the model is closely aligned with your specific domain, providing more relevant and accurate outputs.
- **Reducing software development costs:** The platform democratizes application development to non-technical users while making the technically savvy users become much more efficient.
- **Supporting comprehensive models:** After the models are trained, Databricks provides a unified service for deploying, governing, and querying these AI models and agents, ensuring they're integrated seamlessly into business applications and workflows.
- **Enhancing data security and governance:** Databricks ensures that all your intellectual property remain within your organization's control, reducing data privacy and compliance risks. Databricks offers strong security and governance because of its native foundation in Unity Catalog. You can easily decide exactly what data is available inside the application.
- **Having complete control:** Maintain IP and ownership over both the models and the data. Databricks enables organizations to use their unique enterprise data to build their own independent generative AI solutions.
- **Future-proofing your AI:** With new models and third-party solutions popping up non-stop, Databricks framework and research team enables businesses to select the most suitable approach for their specific use cases and evolve as their requirements evolve.

In this section, you discover a few of the ways to develop AI applications.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Crafting AI applications with Agent Bricks

To build and deploy an effective AI agent, an AI agent system is essential, whether it involves a single agent or multiple interacting agents. AI agents are intelligent applications designed to automate tasks and enhance human productivity. These agents can analyze information, make decisions, and take actions to achieve specific goals, freeing up time and resources for strategic initiatives.

REMEMBER

Databricks Agent Bricks is a framework developed by Databricks that simplifies the creation and optimization of AI agent systems tailored to specific use cases, leveraging your own enterprise data for improved performance. Agent Bricks gives you a full suite of tools to build, tune, evaluate, and deploy high-quality AI. You can quickly create and deploy AI agent workflows in Databricks through reusable, composable building blocks.

Agent Bricks lets you

1. **Build any agent in any way.**

   Custom code agents let you use any framework and model. *Declarative agents* let you build agents with natural language prompts. *AI functions* add intelligence directly in SQL.

2. **Evaluate and optimize quality, measuring every interaction.**

   You can optimize accuracy, latency, and cost and continuously improve with built-in judges.

3. **Govern and scale securely, scaling agents safely across teams and environments.**

   With Unity Catalog, your agents get centralized access control and audits and can govern your data, models, and external tools, too.

Start with the business problem you're trying to solve. That can be a single agent or a multi-agent system for more complex applications. The list of use cases is ever-growing, but Agent Bricks can help with the following example agents:

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

- **Information extraction agents** turn unstructured text like PDFs and images into structured fields such as names, dates, and entities.
- **Knowledge assistant agents** build really high-quality and contextually rich responses with retrieval augmented generation (RAG), grounded in specific, authoritative sources. It delivers fast, accurate answers grounded in your data.
- **Supervisors of multiple agents** provide highly adaptable AI solutions capable of handling intricate workflows in areas like supply chain optimization or multichannel customer engagement. In these systems, agents that are designed for specific functions — such as planners, task executors and data retrieval specialists — are orchestrated together.
- **Custom LLM agents** translate text and provide sentiment analysis with an understanding of your own business.
- **Customer support automation agents** are intelligent chatbots that handle routine inquiries, provide accurate information with proper citations, and escalate complex issues when necessary.
- **Sales and marketing agents** automate lead qualification, generate personalized communications, and analyze customer data for campaign optimization.
- **Data analysis and reporting agents** collect, process, and analyze data from multiple sources, generating insights and reports without manual intervention.
- **Task automation agents** streamline scheduling, inventory management, order processing, and workflow coordination across operations.
- **Classical AI agents** are used for predictions and classifications, such as fraud detection, demand forecasting, and churn predictions.

Databricks makes creating agents easy, as shown in Figure 4-1.

You define the high-level outcome with a prompt, and Databricks creates LLM judges and evaluations, optimizes with the latest models and techniques, and helps you choose the right balance of quality and cost for your use case. Databricks helps with agent learning based on human feedback. You easily and continuously repeat this process to keep improving over time.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

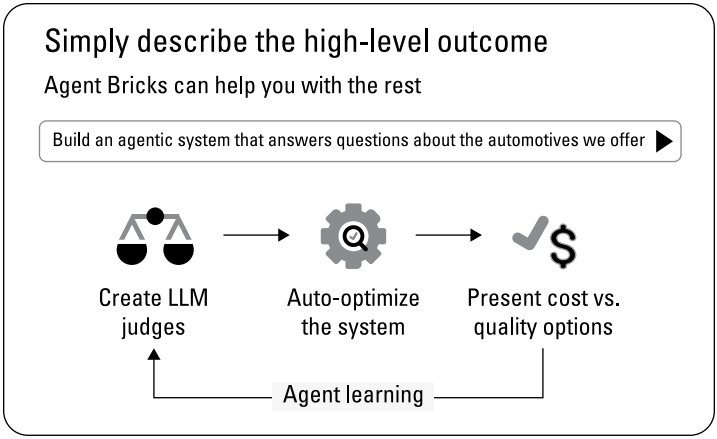


FIGURE 4-1: Agent Bricks helps you build high-quality agents on your own data.

For example, you can build a recommender system: If a bookstore is out of stock of a book, you don't want the customer to walk away, so you want your salesperson to recommend the next-best book to the customer. Agent Bricks can go beyond a traditional recommender system, for example, creating a prompt such as, "Don't recommend the book's sequel if they didn't read part 1."

WARNING

For agent connectivity, the big roadblock has been data silos. Every tool — Salesforce, Slack, Microsoft Teams, GitHub, Google Drive — requires a custom connector for every agent framework. Model context protocol (MCP) is the new open standard that solves this, with its universal way to connect agents to data and tools. These can be built-in tools, third-party tools, or your own custom proprietary internal application programming interfaces (APIs).

Self-serve insights with AI/BI

Companies want to use data intelligence to then extend business intelligence (BI) to every team. After you have great data, you need to get value out of it. This is where AI/BI comes in to marry BI with the power of data intelligence. AI/BI offers two main capabilities:

- **AI/BI Genie:** Simply talk with your own data in the context of your business. Find out more in Chapter 3, and at `www.databricks.com/product/business-intelligence/ai-bi-genie`.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

- **AI/BI Dashboards:** Go beyond traditional BI tools. Let business users self-serve their own AI-driven insights without waiting on technical BI teams. Find out more in Chapter 3, and at `www.databricks.com/product/business-intelligence/ai-bi-dashboards`.

Databricks Apps

Databricks Apps is a fast and secure way to build data and AI applications on the Databricks Data Intelligence Platform. It's really about a broader architectural shift of moving apps to where the data and AI reside instead of the old way of moving the data and AI to the app.

Databricks Apps integrates with all Databricks' offerings, helping you build simple yet powerful apps that all run on serverless compute. Databricks Apps is powered by data intelligence, built with one-click integration with Lakebase as the transactional engine. This feature is secure and governed out-of-the-box with resource-level governance through Unity Catalog. It's 100 percent built on open source and works with a robust open ecosystem, including all popular Python frameworks, such as React, Javascript, Shiny, and more. All of the major vibe coding tools in the ecosystem integrate with Databricks for even faster coding and development.

Putting It All Together

Using the Databricks Data Intelligence Platform greatly accelerates your data and AI success. It simplifies the development of enterprise AI applications has the Databricks data lakehouse as its foundation, and makes it easy for enterprises to create AI applications that understand their data.

The Databricks Data Intelligence Platform helps organizations do the following:

- **Reduce costs and consolidate data:** Breaking down silos frees up budget to accelerate innovation.
- **Enhance and simplify governance and security:** A unified model ensures high-quality, compliant data.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

- **Operationalize AI:** Removing technical barriers enables businesses to leverage AI effectively, to achieve scalable, data-driven innovation across every department.

Wherever you are on the maturity curve, Databricks helps you add significant value by simplifying and democratizing data and AI across your entire organization.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

IN THIS CHAPTER

- Benefitting from a unified data platform
- Uncovering data insights
- Owning your own data and intellectual property
- Saving money

Chapter 5
Ten Reasons Why You Need a Data Intelligence Platform

Businesses collect huge amounts of data from various sources every day. However, having access to vast quantities of data isn't enough; you need a powerful tool to employ the full potential of your data assets. You need a data intelligence platform now because it

- **Eliminates data and artificial intelligence (AI) silos:** This central place serves as one location to unify all your data types and workflows from data warehouses and data lakes to business intelligence (BI), online transaction processing (OLTP), machine learning (ML), and generative AI (GenAI) stacks.
- **Enhances data and AI security:** With a data intelligence platform, you get enhanced data and AI security. This platform offers robust security features that enables your organization to protect sensitive data and meet compliance requirements.
- **Accelerates application and agent development:** Modern businesses need intelligent, data-driven applications, and a data intelligence platform enables teams to rapidly build,

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

deploy, and govern AI-powered applications and composable agents directly on your own enterprise data.

- **Democratizes data and AI across your organization:** Legacy systems rely heavily on technical specialists to get work done. Data intelligence platforms enable business users, analysts, and executives to discover, query, and visualize data by using natural language chatbots and self-service dashboards — without sacrificing governance.
- **Finds more intelligent data-driven insights.** A data intelligence platform better understands the enterprise context of your data, enabling your organization to uncover insights and trends hidden in your data. The platform learns from your organization's data, metadata, usage patterns, and feedback, enabling AI agents and assistants that understand company-specific definitions, rules, and terminology.
- **Accelerates your data tasks through automation and AI assistance.** You work within a single platform that speeds up tasks such as data engineering, ML, and AI, enabling seamless collaboration and efficient workflows. The data intelligence platform automates data pipelines and infrastructure management, reducing manual effort, minimizing errors, and improving reliability.
- **Offers better collaboration:** The platform facilitates easy partnerships among teams and users, helping them share insights, code, and results. This teamwork accelerates data-driven decision-making, which greatly benefits your business.
- **Massively scales, cost-effectively:** The platform handles large-scale data processing by separating storage and compute, allowing your organization to work more efficiently and cost-effectively with its huge amount of data. You get structured, semi-structured, and unstructured data all in one platform — way better than rigid legacy systems.
- **Establishes a single, trusted source of truth:** Siloed data, inconsistent definitions, duplicated datasets, and fragmented metadata undermine trust in data. A data intelligence platform centralizes business semantics, metrics, lineage, and metadata so everyone works from consistent, governed data.
- **Improves return on investment (ROI):** Everyone loves cost savings, right? By consolidating data management and analytics into a unified data intelligence platform, your organization can reduce costs and shift from reactive reporting to proactive, AI-driven decision-making.

These materials are © 2026 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Maximize your company's potential for data+AI

Maximize your organization's potential for data and AI with data intelligence. The Databricks Data Intelligence Platform is built on an open, unified, and governed lakehouse architecture and powered by a data intelligence engine. By using AI, the platform can reason on your own enterprise data, enabling you to harness the full value of your unique data assets. Whether it's ETL, data warehousing, OLTP, BI, traditional or generative AI, data intelligence streamlines and accelerates your path to data-driven success.

Inside...

- The value of data intelligence
- The power and potential of AI
- Features of data intelligence platforms
- Building AI applications
- Why you need a data intelligence platform

databricks

Ari Kaplan is Databricks' Head of Technical Evangelism. He's Caltech's "Alumni of the Decade" and has created the Chicago Cubs' & Baltimore Orioles' analytics departments. **Amit Kara** is Director of Technical Marketing at Databricks and an expert in data management, helping organizations unlock their data's potential.

Go to Dummies.com™
for videos, step-by-step photos, how-to articles, or to shop!

ISBN: 978-1-394-39641-2
Not For Resale

for dummies®
A Wiley Brand

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.