

# Appendix

# Customer

# Outcomes

## Financial Services Industry

---





# Banking & Payments





# Accelerating Petabyte-Scale Feature Engineering with Databricks

## Challenge

- Legacy compute infrastructure frequently timed out on petabyte-scale datasets
- High processing costs for 160+ machine learning features spanning 10+ years
- Fragmented pipelines slowed development and delayed insights

## Solution

- Replaced legacy compute with Databricks Photon for high-performance processing
- Unified historical and real-time feature pipelines into a single platform
- Re-architected and optimized 160+ complex feature transformations

## Impact

- 60x faster compute performance, enabling real-time iteration
- 80% reduction in operational costs
- 7.5B records processed daily
- 6T+ records analyzed historically
- Standardized feature engineering across the enterprise with Databricks

# Vantage Bank Texas Transformed Lending Workflows

## Challenge

- Manual loan reviews slowed approvals and risked errors.
- Policy checks ate into lender time and customer focus.

## Solution

- AI assistant built on Databricks by Cavallo and Vantage.
- Delivered MVP Day 1 for instant impact and iteration.
- Answers complex questions with sources and collects feedback.
- Auto-updates policy knowledge for compliance at scale.
- Slashes costs by scaling down after hours.
- Gives leadership live dashboards for top-down insight

## Impact

- 94% accuracy on tough lending questions.
- Faster, more precise loan approvals.
- Transformed workflows and boosted user trust.
- Framework now powers service, risk, and compliance.



**TECHCOMBANK**

## Ushering personalized banking for millions of customers

### Challenge

Techcombank's data silos and governance gaps slowed innovation, collaboration, and customer-centric banking for millions of users.

### Solution

Unified 50+ systems and 315 branches on the Databricks Data Intelligence Platform—leveraging Delta Lake, Databricks SQL, and Unity Catalog to centralize analytics, democratize AI, and power real-time personalized banking.

[Read the case study >](#)

©2025 Databricks Inc. — All rights reserved

### Impact

**1,000+**

active users accessing data for analytics and reporting

**15.3M**

banking customers supported by personalized AI-driven services

**100+**

ML models and 7,500+ features powering advanced risk and customer analytics





## Making banking more accessible and secure

### Challenge

Maya's fragmented legacy infrastructure made it difficult to scale AI, ensure real-time data access, and drive financial inclusion for millions of Filipinos.

### Solution

Modernized to a unified Databricks Lakehouse for centralized analytics, reliable ML lifecycle management, and scalable, governed data—fuelling rapid development in lending, fraud prevention, and user engagement.

[Read the case study >](#)

©2025 Databricks Inc. — All rights reserved

### Impact

**\$2.1 billion**

in loans disbursed by ML-powered underwriting

**Up to 98%**

fraud loss reduction using AI transaction pattern recognition

**45%**

increase in ARPU from hyper-personalized AI-driven recommendations





## Challenge

- Top-15 global bank, struggled with siloed data platforms and legacy systems across regions
- These silos led to inconsistent data, high costs, compliance risks, and difficulty meeting regulations like BCBS239

## Solution

- Modern Data Platform: Partnered with Deloitte to build a GenAI-powered platform on Azure Databricks
- Unified Architecture: Lakehouse model integrated loans, deposits, securities, and more
- Global Sharing: Replaced file transfers with Delta Sharing for secure, real-time data exchange
- Built-in Governance: Unity Catalog and Collibra ensured access control and data lineage
- GenAI Automation: Accelerated ETL and data quality for faster, more reliable onboarding
- Federated Model: Business units own their data products with central governance
- Quick Wins: Delivered an 8-week MVP to prove value and gain internal buy-in

## Impact

- Migrating 20+ banking systems and all structured data to one platform within a year
- Faster delivery of analytics and compliance tools
- Better data quality and trust for regulatory reporting
- Eliminated manual and Excel-based processes
- Freed up teams to focus on insights over data wrangling
- Set up for global data sharing and natural language querying by 2027

# Optimizing Azure Databricks for security and cost efficiency

## Challenge

- Migrate from on-prem analytics to a secure cloud platform in a regulated finance environment.

### Pain Points:

- Complex network caused high costs and slow cluster startups
- GDPR required strict data separation and governance
- Storage was bloated

## Solution

- Cloud Migration: Phased move to Azure Databricks with Terraform automation
- Smarter Architecture:
  - Unified workspaces in one VNet to cut traffic
  - Single-VNet per subscription to avoid peering costs
  - Serverless compute and egress policies for security and savings
- Efficient Sharing: Used views instead of duplicating data
- Built-in Guardrails: DLP, download blocks, and Azure security integration
- Governed Access: Unity Catalog enabled federated ownership with centralized standards

## Impact

- 60% lower networking costs
- 30% lower compute costs
- Faster, reliable cluster startups—no more timeouts
- Reduced storage from 1 PB to actual usage—no duplicates
- 120+ users and key use cases (AML, GenAI bots)
- GDPR compliant and strong data governance with Unity Catalog
- Streamlined operations and improved efficiency



# Block improves employee productivity and data access with Claude in Databricks

## Challenge

Payment processing company Block had vast internal data but lacked a scalable way to make it accessible to non-technical employees across departments. This limited productivity and slowed innovation, as teams depended on specialists for insights, coding, and system navigation.

[Watch the video >](#)

## Solution

By integrating Claude with the Databricks Data Intelligence Platform through their internal agent goose, Block created an AI-powered system that automates workflows, generates SQL, and connects to internal tools. This democratized data access, enabling thousands of employees to analyze data and build solutions without needing deep technical skills.

## Impact

**30%**

Expected employee time savings with AI

**75%**

Of engineers saving 8 to 10+ hours every week

**\$10M**

Additional productivity gain vs. forecast, with 3x better price/ performance for serving models

Delta Live Tables drives real-time omnichannel app monitoring for Navy Federal.

## Challenge

Navy Federal's data pipelines ran daily, which was too slow for their critical user migration project. They faced tight deadlines and needed to manage billions of events while scaling their infrastructure for real-time insights. This posed significant challenges in delivering timely, actionable data.

## Solution

Delta Live Tables (DLT) enabled Navy Federal to create a near real-time omnichannel app monitoring dashboard with latency reduced to under 10 minutes. DLT efficiently managed billions of events, ensuring robust performance and reliable insights.

## Impact

# 9 Billion

Events streamlined for 9 months, 24/7, with ~100% zero maintenance

# 6 Weeks

To build a new application into production with DLT



## About

Founded in 2002, Zafin is a global provider of SaaS solutions for core banking modernization and transformation.

## Key Metrics

**30-40%**

Reduced time to market

**50-70%**

Lower cloud costs for analytics

**10X**

Faster computation speed

# Empowering Personalized Banking Experiences

## The Challenge

Zafin assists banks in determining customer relationship tiers and providing personalized rewards. To achieve this, Zafin processes substantial data volumes—100 million accounts with over 10 billion balances. Using a traditional tech stack to build and scale this solution proved inefficient and unsustainable at such a scale.

## Solution

With the Data Intelligence Platform, Zafin leverages Unity Catalog for robust data governance to isolated single-tenant customer environments and maintain compliance with financial services regulations. Applying Databricks SQL and Notebooks, it built a data-intensive tiering application that evaluates customer balances, product holdings, and transaction patterns to determine relationship tiers. This allows banks to deliver personalized rewards and benefits to customers while nudging customers toward the next tier with targeted engagement, increasing retention and loyalty. And with Databricks Serverless capabilities, they save costs for analytics while providing a seamless, responsive experience.

## Outcomes

The unified environment of Databricks improved productivity and collaboration, enabling Zafin to deliver personalized rewards, actionable business insights, cost-efficient data transformations, and robust data governance, significantly enhancing the value provided to their banking customers.



# Mastercard Scales GenAI Innovation

## Challenge

Mastercard struggled to scale over 300 GenAI proof-of-concept (PoC) projects due to concerns over cost, governance, and risk, leaving many promising ideas stuck at the demo stage. They needed a secure and unified platform to operationalize these use cases across the organization.

## Solution

Using Databricks, Mastercard transformed their initial "Product Onboarding Assistant" PoC into the "Mastercard Assistant" platform, which efficiently supports multiple GenAI applications. This enables scalable and secure deployment of GenAI, driving innovation while managing governance and costs.

## Impact

**300+**

GenAI applications deployed on Mastered Assistant

**50%**

Reaching goal to have more than half of their revenue stem from data products and services



# Community bank masters modern data platform for exponential growth

## Challenge

Coastal Bank's vision was to implement Banking as a Service, necessitating real-time data exchange with partners. They needed a scalable solution that ensures security, privacy, and compliance, all managed by a small team.

## Solution

With the Data Intelligence Platform, Coastal overcame issues from a previous data lake system. This platform enabled secure live data sharing, supported quick growth through partner acquisition, and increased the speed of data integration from partners. This led to a dramatic increase in the number of customers served, from 40,000 to 6 million.

## Impact

**<10 mins**

To securely share large datasets across organizations

**99%**

Decrease in processing duration (2+ days to 30 min.)

**12x**

Faster partner onboarding by eliminating sharing complexity



# Databricks Fuels NAB's Data-Driven Transformation for 8.5 Million Customers

## Challenge

NAB, Australia's largest business bank, needed to replace slow and costly legacy data platforms to enable real-time analytics, AI, and regulatory compliance for millions of customers.

## Solution

NAB built a cloud-native data platform on Databricks, unifying engineering, analytics, and AI workloads on a modern, governed lakehouse.

## Impact

**50%**

Cut data ingestion costs

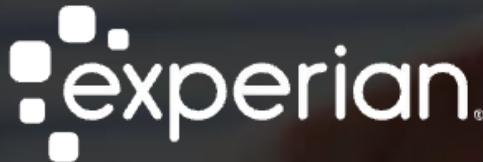
**30%**

Boost in ML model and query performance

**1,000+**

users enabled with self-service, governed data access





# Delivering faster, more helpful customer service

## Challenge

Experian aimed to scale GenAI for customer support by automating email responses.

- Existing infrastructure lacked native LLMOps and made fine-tuning slow and costly (86+ hours).
- Cloud setup lacked governance, compliance traceability & flexibility to test open-source models.
- Difficulty retrieving relevant content for AI-generated responses.

## Solution

Databricks Platform let them build, fine-tune, and govern their GenAI email assistant, Latte, at scale.

- **Mosaic AI:** Fine-tuned Llama 8B model to power Latte and automate 1,000+ emails/day.
- **DBRX LLM:** Generated richer instruction data to improve accuracy and flexibility.
- **Vector Search:** Real-time retrieval of relevant content for intent-aware responses.
- **Unity Catalog + MLflow:** Full lifecycle governance, versioning & compliance.

## Impact

**11x**

Faster fine-tuning time for Llama 8B model

**35%**

Of email responses to customers are now automated

**8%**

Increase in NPS, signaling stronger customer satisfaction and loyalty



# Experian uses the Databricks to fight identity theft

## Challenge

Experian was responsible for managing vast data assets, including structured, unstructured data, and machine learning models. They needed a unified platform to improve data governance and team collaboration across these assets.

## Solution

By adopting the Databricks Data Intelligence Platform with Unity Catalog and Vector Search, Experian streamlined data management and governance. The platform allowed for easier machine learning asset management and improved collaboration among engineers.

## Impact

Experian successfully launched their ID health score product, helping consumers assess their identity theft risk using machine learning and dark web data. The improved data foundation enhanced product offerings and consumer protection against fraud.

[Watch the video >](#)



HSBC

# Boosts Mobile App Experience with Databricks

## Challenge

As one of the largest global banks, HSBC wanted to personalized mobile banking and modernize product development by leveraging data, AI and ML. HSBC built **PayMe**, a social app that facilitates cashless transactions between consumers and their friends instantly and securely.

## Solution

Databricks provided a unified platform for data science, engineering and business analysts to build new data products on a shared, single source of data.

## Impact

**4.5x**

Improved engagement in HSBC's mobile app

**6 seconds**

Data processing drastically improved from 6 secs from 6hrs

**1 data lake**

Replacing 14+ databases to improve productivity

ABN AMRO uses Databricks to deliver financial services transformation at a global scale, delivering automation and insight across operations.

## Use Cases

- Personalized finance
- Risk management
- Fraud detection
- + many more

## Challenges

- Limitations associated with legacy Hadoop systems; disparate data difficult to access and unify for analytics; siloed teams

## Solution

- Delta Lake brings reliability and performance to data lake environment
- Able to collaboratively explore, create, and unify datasets at scale with SQL analytics to reduce time-to-insight
- Interactive notebooks & ML flow improves data science productivity across the ML lifecycle

## Impact

**10x**

faster time-to-market of new ML solutions vs. legacy infrastructure

**50+**

use cases delivered with 100+ more in the coming year



# Uses Databricks to Facilitate Thousands of Small Business Loans

## Challenge

With 100M+ customers, Intuit needed to leverage data and AI to unlock insights and personalize customer experiences, such as guided document navigation, cash flow forecasting, fraud prevention, and connecting customers with the right support agents.

## Solution

Databricks Lakehouse extracts meaningful cross-company insights for business decision making and feeds ML models designed to simplify the loan application experience

## Impact

**2M**

ML models in production

**\$1.4B**

in loans facilitated to 40,000 small businesses

**20x**

faster time-to-insight

Banco Hipotecario uses Databricks to increase customer acquisition and cross-sells while lowering the cost for acquisition, greatly impacting overall customer retention and profitability.

## Use Cases

- Personalized banking
- Legacy SAS system was rigid and complex to scale
- Data was disorganized and inaccurate
- Workflows were time - consuming, impacting ability to perform timely analytics
- Poor data science collaboration

## Why Databricks?

- Delta Lake brings reliability and performance to data lake environment
- Able to collaboratively explore, create, and unify datasets at scale with SQL analytics to reduce time-to-insight
- Interactive notebooks & ML flow improves data science productivity across the ML lifecycle

## Impact

### Reliability

Data Lake reliability accelerates ETL for analytics and ML, increasing customer engagement and lifetime value lifetime value

**90%**

increased fin cross-sells of new products, improving customer retention and lifetime value

**35%**

reduction in cost of new customer acquisition

## Challenge

IFC struggled with manual processing of vast amounts of unstructured ESG data, causing lengthy analysis periods and suboptimal model performance. Limited collaboration tools hindered team efficiency and scalability.

## Solution

IFC implemented the Databricks Data Intelligence Platform, leveraging GPU computing and lakehouse architecture to streamline data processing, optimize model performance, and improve collaboration. They developed MALENA, an AI-powered platform, to transform analysis for emerging market investors.

## Impact

**19K**

Sentences analyzed per minute with MALENA, built with Databricks

**950x**

Increase in analysis time

**1.2K**

ESG risk terms rapidly identified

# Capital Markets



NEW



Brewin  
Dolphin

# Helping Financial Planners gather review insights for client meetings

## Challenge

RBC Brewin Dolphin needed to modernize preparation for annual Client Advice Reviews. Data was scattered across multiple systems, and assembling meeting packs was manual, time-consuming and prone to errors—reducing advisor focus on client relationships.

[Read the case study >](#)

## Solution

Using the Databricks Data Intelligence Platform, RBC Brewin Dolphin centralized structured and unstructured data with Delta Lake and controlled access via Unity Catalog. Lakeflow Jobs orchestrates GenAI pipelines to extract insights from documents and generate meeting packs.

## Impact

**80-90%**

Of meeting pack preparation automated with GenAI

**~50%**

Estimated reduction in administrative costs in some RBC BD offices

**4,700**

Hours saved annually due to reduced preparation time



# RBC uses AI to transform capital markets research

## Challenge

RBC Capital Markets faced mounting pressure to deliver expert research and insights across a broader coverage universe, faster than ever before, as clients demanded near real-time, differentiated commentary. Traditional analyst workflows were highly manual and slow, with post-earnings research notes taking up to 45 minutes per company—creating bottlenecks and risking missed market opportunities.

## Solution

RBC partnered with Databricks to build Aiden QuickTakes, a proprietary AI application embedded within a secure, cloud-native Data Intelligence Platform. This unified environment—leveraging Databricks' Unity Catalog for governance and Mosaic AI for end-to-end AI capabilities—automates data ingestion, analysis, and content generation, enabling analysts to produce high-quality research drafts in minutes while maintaining strict security and compliance standards.

## Impact

**60%**

Reduction in research turnaround time

**Thousands**

Of hours reclaimed for higher value work

**Faster insights**



# How State Street Unifies Data and Drives Innovation with Databricks

## Challenge

State Street aimed to unlock the potential of structured and unstructured data for faster, smarter decision-making in the financial services sector. Unfortunately, they dealt with:

- Difficulty accessing insights from siloed data, including unstructured formats like PDFs.
- Fragmented data sources hindered seamless integration and timely insights.
- Inefficient workflows to transition from raw data to actionable AI outcomes.

[Watch the video >](#)

©2025 Databricks Inc. — All rights reserved

## Solution

Databricks Data Intelligence Platform empowered State Street to unify its data and AI workflows for efficient, scalable insights.

- **Unity Catalog:** Centralized governance and seamless indexing across structured and unstructured data.
- **AI Workflows:** Enabled large-scale data processing and curation for actionable insights.
- **Unified Platform:** Integrated structured and unstructured data for streamlined AI-driven decision-making.

## Impact

## Unified

data accessibility, unlocking actionable insights from both structured & unstructured data

## Faster

and smarter decision-making after breaking down silos, enabling timely insights in a competitive financial market



# Unlocking Financial Insights with a Custom Text-to-SQL Application

## Challenge

The Intercontinental Exchange (ICE) faced the challenge of enabling non-tech users to access structured data without needing to understand complex data models, schemas, or SQL queries. This hindered their ability to efficiently extract insights from large volumes of financial data.

## Solution

By collaborating with Databricks, ICE (part of the NYSE) implemented a RAG pipeline using tools like Unity Catalog and Vector Search. This text-to-SQL app allowed users to generate accurate SQL queries from natural language inputs, enhancing data accessibility and operational efficiency.

## Impact

**77%**

Syntactic accuracy

**96%**

Execution match

[Read the blog >](#)

# Goldman Sachs

## Challenge

Goldman Sachs wanted the Apple Card to reach as many customers as possible — without significantly increasing risk. But their data architecture could not easily support the real-time machine learning required to make it successful.

## Solution

Using Databricks, Goldman Sachs deployed a Lakehouse that processes 30TB a day across a large portfolio of data providers to accurately predict constantly evolving lender risk profiles.

## Impact

**\$50M**

in revenue from improved credit risk approval models

**\$53M**

in revenue from better cross-selling promotions

# GlobalLINK

by  
State Street

## Challenge

GlobalLINK, State Street's trading tech division, needed to deliver faster insights and AI-driven decision-making across institutional platforms. But legacy on-prem systems couldn't scale to process over a billion tick updates daily or support strict data boundaries. Critical analytics like spread performance took hours to run, slowing response times and limiting innovation across trading, compliance, and liquidity workflows.

## Solution

The Databricks Data Intelligence Platform unified GlobalLINK's data foundation, enabling real-time analytics, secure multi-tenant access, and automated ML development. By modernizing its architecture, the team eliminated infrastructure overhead and now powers faster spread analysis, intelligent liquidity optimization, and new monetizable data products at scale.

## Impact

**1B**

tick updates stored and processed daily

**98%**

decrease in pricing spread analysis—now runs in 4 minutes instead of 4 hours

# Insurance

---



## Challenge

Suncorp faced difficulties in providing real-time, accurate responses to insurance agents while managing scalability, performance, and the high costs of deploying generative AI models. Balancing these without sacrificing response speed or accuracy proved to be a significant challenge.

## Solution

Suncorp implemented Databricks Mosaic AI to manage and deploy multiple LLMs like ChatGPT, integrating real-time data orchestration. This solution optimized the balance between accuracy, cost, and scalability, improving operational efficiency and customer service interactions.

## Impact

**1K**

Employees leveraging “single view of claim” app within home insurance teams

[Read the article >](#)



With Databricks, AXA France has fundamentally changed how they operate and use data – unifying 200TB of data from 54 data sources into a single enterprise data platform.

## Challenge

With over 6.3 million customers, AXA France's drive to provide innovative, personalized and differentiated financial services has ushered in a massive digital transformation as a top company objective.

## Solution

With Databricks, AXA France unified 200TB of data from 54 data sources into a single enterprise data platform. A dynamic data strategy has also helped AXA democratize data and empower end-users to be more productive, and not just the technologists.

## Impact

**50%**

Reduction of TCO

**200TB**

Unified in a single data Databricks Intelligence platform, across 54 data sources

**20x**

more users now have easy access to data



## Challenge

Legacy systems were unable to deliver personalization at scale for 10,000 advisors and for millions of customers. This created data latency issues, increased the cost of data management and produced unwarranted security constraints.

## Solution

Databricks provides an integrated data analytics solution built on an open lakehouse architecture to power their Insights Platform, which provides wealth management insights from aggregating transactional and behavioral data. This allows NWM to create a digitally-enabled, automated, end-to-end underwriting experience.

## Impact

**3.5x**

Accelerate batch cycles from 7 hours down to 2 hours

**50–75%**

faster time to market, from 4–6 weeks to 1–2 weeks

**10,000**

users performing ad hoc analysis with Databricks SQL



## Challenges

With hundreds of millions of claim records to process for insurance pricing, Nationwide's legacy infrastructure (SAS) struggled with providing accurate insights to predict the frequency and severity of claims.

They needed to modernize their infrastructure to scale the use of data, analytics and AI.

## Solution

Moving to Databricks unified data platform simplifies infrastructure management, enabling fast data pipelines at scale, and enabling faster time-to-market for ML solutions. Embed deep learning to provide more accurate pricing predictions, resulting in more revenue from claims.

## Impact

**9X**

Performance gain with data processing at scale, with improved runtime of their entire data pipeline

**50%**

Faster AI model training

**60x**

Accelerated model scoring from 3 hours to less than 5 minutes, a 60x improvement



## Challenge

Allianz Direct struggled with inefficiencies in their contact center operations, where agents were burdened by mundane tasks that hindered effective customer engagement. They needed a solution to streamline these processes while ensuring compliance with corporate governance and AI ethics.

## Solution

They implemented Databricks Mosaic AI to develop an agent-assist copilot that intelligently provided accurate information in real-time, significantly improving operational efficiency. This seamless integration allowed for quick execution and has inspired further innovation across their contact centers.

## Impact

**15%**

Improved answer accuracy provided to agents with copilot

[Read the case study >](#)



## Challenge

Santalucía Seguros had difficulty managing and accessing varied documentation, hindering their ability to provide timely and accurate responses to customer inquiries.

[Read the blog >](#)

## Solution

They built an accurate Virtual Assistant using RAG on their Advanced Analytics Platform powered by Databricks. Mosaic AI Model Serving facilitated seamless use of external LLMs, managing configurations, credentials, and security access through REST API.

## Impact

**24/7**

Customer support availability

**Seconds**

To answer customer queries with the help of the Virtual Assistant



*First American*

## Challenge

First American faced challenges extracting data from millions of historical title policies stored as image files with minimal metadata. They needed to improve extraction speed and accuracy while working with limited GPU resources and a tight timeline.

## Solution

First American DNA leveraged Databricks Mosaic AI to fine-tune open-source LLMs like Mistral and Llama, using Delta Lake for efficient data storage. Databricks' serverless compute and Model Serving helped scale the solution without extensive hardware management, reducing costs and time.

## Impact

**70%**

Reduction in cost

**75%**

Accelerated project timelines, reducing from months to weeks

**10x**

Extracted more data fields, turning documents into actionable insights



## Challenge

CZ had significant data fragmentation, with structured, semi-structured, and unstructured data siloed across the org. It was difficult for data analysts and scientists to access and apply analytics effectively. This limited their ability to innovate, scale AI solutions, and maintain a consistent, governed data landscape.

## Solution

By adopting the Databricks Data Intelligence Platform, CZ unified its data infrastructure, enabling seamless collaboration across teams using multiple programming languages. The integration of Unity Catalog and Medallion Architecture provided scalable governance, organization, and lifecycle management of data products.

## Impact

Enhanced data maintainability & reduced reliance on call centers.

Accelerated premium pricing models.

Improved CX with AI mobile app.

Automated quality checks on claims and call summaries.

# Financial Data Providers



# Unifying Financial Data Platforms with Databricks

## Challenge

- Legacy systems and siloed workflows slowed product launches
- Manual handoffs and fragmented toolchains caused inefficiency
- High DevOps overhead from maintaining multiple technologies
- Needed to support diverse, high-volume data products with strong governance

## Solution

- Unified all data workflows on Databricks Lakehouse
- Centralized data mastering, analytics, and governance
- Standardized pipelines and toolchains for all teams
- Enabled rapid onboarding, streaming/batch data, and open standards

## Impact

- 100B+ transactions processed daily to power real-time insights at global scale
- Product launches 4x faster — from months to weeks or days
- 75% faster model run times — critical analytics in 10 minutes, not 60
- Thousands of data products managed on one unified platform



## Challenge

FactSet's previous solution, Synapse ETL, was inefficient from 2 perspectives: performance and complexity. First, Synapse ETL pipelines were slow, each taking over 30 minutes to run. Second, Synapse's architectural complexity required stitching together various tools and limited data engineering format choices. Security also became a big challenge.

## Solution

Databricks Data Intelligence Platform allowed FactSet to unify their data use cases and scale in a cost-efficient manner. The Databricks Data Intelligence Platform unifies data lake assets with warehouse-level performance, including unified governance tools like Unity Catalog and scalable tools like open-source Delta Lake and Apache Spark.

## Impact

**80%**

Reduction of ETL costs aided by faster processing times, and auto scaling clusters

**90x**

performance improvement



## Challenge

FactSet is a leading provider of financial data and analytics software. They tried deploying a leading commercial LLM to do natural language queries but it was not accurate and too slow.

[Read the blog >](#)

## Solution

FactSet built a Compound AI system using Databricks. In addition to using commercial LLMs, they fine-tuned an OSS model with Mosaic AI Model Training and deployed a RAG solution with Vector Search.

## Impact

**42%**

More accurate responses (to 85% accuracy)

**2.5x faster**

Better e2e latency

# S&P Global

## Challenge

S&P's merger with IHS Markit created infrastructure complexities that hamstrung their ability to democratize data and insights. Their legacy data warehouse also struggled with unifying alternative data sources across the business, including credit risk, ESG and Marketplace Workbench.

## Solution

Databricks Data Intelligence Platform allows S&P to easily process 700 billion data points from a centralized platform. With data and AI at their fingertips, S&P is able to unlock ESG insights, mitigate credit risk and share actionable data with customers and partners.

## Impact

**700B+**

data points from a centralized platform

**65%**

improvement in cost efficiency from retiring legacy infrastructure

**950+**

users of the platform, up from only 3 users in 2018



## Challenge

ADP Assist is a human capital management tool that combines ADP's dataset with AI to help users with HR tasks. Their primary challenge is having their application provide the right answers about payroll and taxes that are accurate and secure. Off-the-shelf LLMs could not reason over their unique HR dataset and were expensive as they scaled the solution.

[Watch the video >](#)

©2024 Databricks Inc. — All rights reserved

## Solution

ADP built OneData, a central data lake that stores all of their data using Delta Lake and OneAI, a AI Agent system built using Mosaic AI. They also used Unity Catalog to oversee governance, security, and observability in a single place. This solution allowed them to deploy and fine-tune open source LLMs that understood their unique HR data at a much lower cost than proprietary models.

## Impact

**5X**

Better price per performance of open-source models

**100**

Products to operationalize gen AI capabilities across

# ESGpedia

## Challenge

ESGpedia sought to scale AI- driven Environmental, Social, and Governance (ESG) data insights more effectively.

- Difficulty unifying disparate data sources for analysis, hindering timely, business-specific insights.
- Hard to manage 300+ data pipelines across multiple platforms.
- Challenges coordinating multiple models for scalable AI workflows.

## Solution

Databricks Data Intelligence Platform enabled ESGpedia to unify data storage and processing and deploy custom AI solutions at scale.

- **Unity Catalog:** End-to-end data and AI governance.
- **Mosaic AI Agent Framework:** Multi-agent system to optimize data classification.
- **Mosaic AI with RAG:** Deliver accurate, context-specific ESG insights for clients.

## Impact

**4x**

Cost savings in data pipeline management

**6 months**

To transition 300 pipeline, accelerating insights delivery

Yipitdata uses Databricks as a central platform to manage all data workflows, saving on operational costs and reducing processing times.

## Use Cases

- Data-driven investments
- Billions of requests collecting data from hundreds of websites; siloed teams; inability to scale their data processing and analytics on legacy data warehousing technology

## Solution

- A unified platform to manage the entire data analytics workflow from data ingestion to downstream analytics.
- Integrated cluster management with autoscaling has simplified infrastructure management while lowering operational costs

## Impact

**60%**  
lower database costs

**\$2.5 million**  
annual savings

**90%**  
reduction in data processing time