eBook

Driving Innovation in Financial Services With a Unified Approach to Data + AI

A guide to data analytics and machine learning for the modern financial institution

databricks

Contents

The New Normal: How Data Is Transforming Financial Services

Data and AI are great democratizing forces in financial services. The synergy between these two forces fuels progress toward the goal of digital transformation that many Financial Services Institutions (FSIs) have today.

To realize the full potential of data and AI, FSIs are rapidly trying to modernize their technology infrastructure — adopting cloud and data analytics technologies to help them transform their massive volumes of data (e.g., financial data, customer data, loan data, transaction data, alternative data, ESG data) and bring agility and innovation to their organizations.

As a result, the traditional advantages that FSIs enjoyed, such as capital and scale, are no longer the main differentiators going forward. The winners of tomorrow will thrive based on their ability to attract engineering talent and unlock the value of data with the right tools and technologies.

The rising importance of AI

THE VISION:

64% of financial institutions surveyed anticipate employing AI in all these categories within the next two years:

- NEW REVENUE STREAMS
- RISK MANAGEMENT
- PROCESS AUTOMATION
- CUSTOMER SERVICE AND CLIENT ACQUISITION

THE REALITY:

Only **16%** of respondents currently employ AI in all these areas

The Opportunity for Data + AI

AI is changing how financial institutions generate and utilize insights from data. Leaders across banking, insurance, capital markets and wealth management are harnessing the power of data and analytics to solve strategic challenges and make smarter decisions that minimize risk, prevent fraudulent behavior, and drive sustainable value creation.

Widespread adoption of AI use cases across the industry

Leading financial services companies are investing billions in technology, moving to the cloud and implementing emerging technologies, reports CIO Dive.

> “In order to continue to deliver the levels of service our customers require, we needed to evolve our business model with data at the forefront.”
>
> **ANURAG SEHGAL**
> Managing Director, Head of Global Markets at Credit Suisse

When it comes to operationalizing data and AI to minimize risk, build customer relationships, and drive higher returns on equity, there are three key themes that should be considered top priority:

RISK MANAGEMENT

Traditional banks relying on on-premises infrastructure can no longer effectively manage risk. Moving to a cloud-based data analytics platform helps FSIs adopt a more agile approach to risk management by unifying data and AI.

DATA-DRIVEN ESG

ESG is widely considered a top-level initiative to achieve resilience and sustainable profitability in a rapidly evolving economic and environmental landscape. This solution provides companies and investors a holistic and data-driven view into ESG performance.

FRAUD PREVENTION

Curbing fraudulent or malicious behavior — from fraudulent securities trading to money laundering — is key to mitigating negative revenue impact. This solution provides ways to leverage data and machine learning to detect anomalies.

With data and AI, FSIs can tap into numerous business use cases across sectors to reduce risk, deliver better customer experiences, and fight fraud.

databricks

The Opportunity for Data + AI

Use cases by financial services sector

BANKING AND PAYMENTS

Deepen customer relationships through secure and personalized services

- Customer 360
- Risk Management
- Regulatory Compliance
- Fraud Detection

INSURANCE

Optimize the entire insurance value chain with automation and predictions

- Claims Automation
- Actuary Science
- Risk Monitoring
- Case Management

CAPITAL MARKETS

Reduce risk while boosting customer ROI with intelligent investments

- Back-Testing
- Risk Management
- Trading Cost Analysis
- Regulatory Reporting

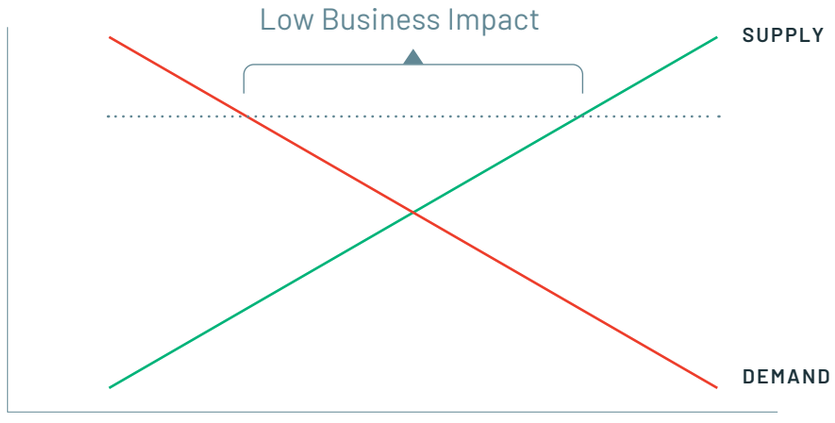
databricks

The Disrupted State of Supply and Demand

While *digital transformation* is a commonly used buzzword, oftentimes, it is just that: a buzzword. The vast majority of AI and data science projects that are supposed to drive digital transformation fail. The key initiative FSIs must execute on is bridging the gap between the supply and demand of data.

Today, there exists a gap between the exponential growth of data (supply) and the ability to use it (demand). As a result, the supply of data typically overshoots the demand for data and the gap between the two prevents financial services institutions from fully monetizing their data assets.

Supply and Demand of Data



The imbalance between data volumes and the ability to leverage it prevents value creation

To effectively leverage data, FSIs first must think of data as an asset rather than a liability. On the supply side, the explosion of data from a broad and highly disjointed data ecosystem has been met with an increasingly robust ability to store it. Cheap and near infinite storage capacity offered by public cloud vendors has enabled FSIs to store more and more data in data lakes. While retention of data may have certain regulatory benefits, data lakes have become dumping grounds or data swamps, as the myriad of data flowing from various data sources becomes unwieldy to manage, let alone the nightmare of extracting actionable insights in a timely manner.

Why Financial Services Institutions Struggle With Data + AI

The foundation for successful deployment of AI is in the data. Without quality data, organizations cannot consume it for business purposes. Analytics fail. Predictive models don't work. Simply put, for most corporations, the inability to access and unify separate data sources to meet the needs of business use cases has been a blocker rather than an enabler of transformative innovation.

Exasperating the complexities of the data is the fact that many traditional banks and FSIs still rely on legacy infrastructure that is too costly and complex to scale. What use is all this data if you aren't able to analyze and leverage it? FSIs must abandon the computational inefficiencies of legacy technologies and build an agile practice in the cloud capable of rapidly responding to market and economic volatility through the use of data and advanced analytics.

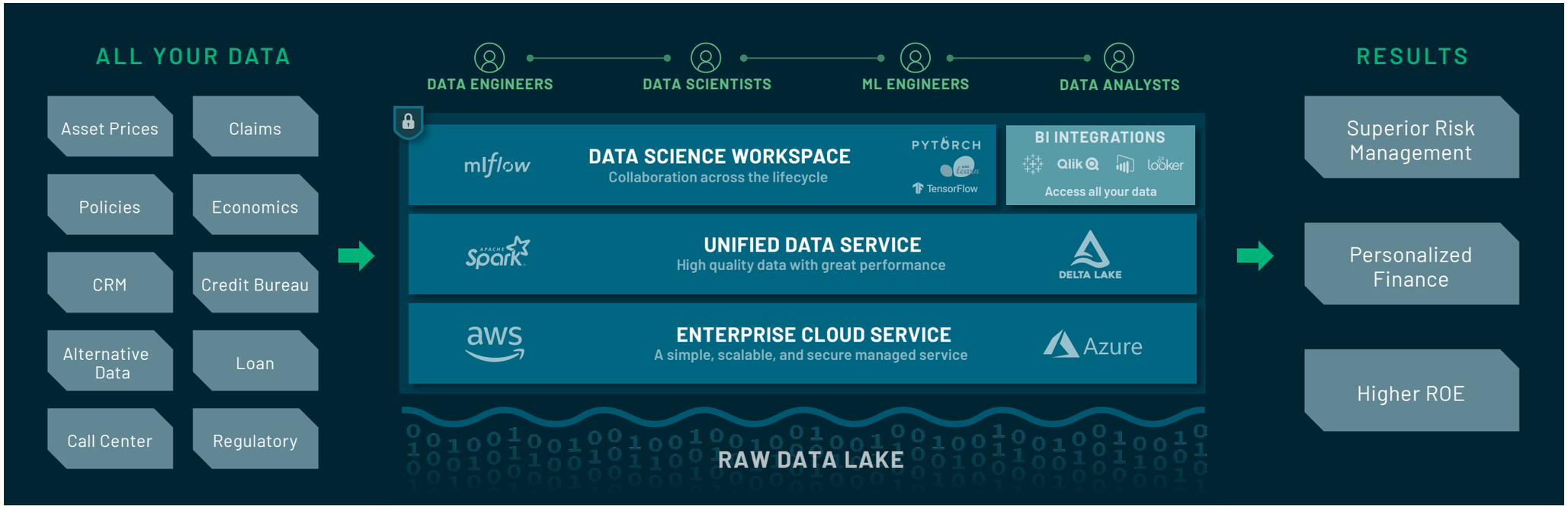
Poor team structure, collaboration and outdated skill sets are slowing progress. Analysts and engineers skilled in legacy systems need to be retrained to leverage the latest data and AI technologies. Along with siloed data, FSIs with traditionally rigid organizations also have siloed teams who don't have the means to efficiently share data, operationalize insights and collaborate on cross-organizational initiatives.

In order to succeed in the journey of digital transformation, FSIs must adapt to today's agile and data-driven times while modernizing their approach to data and analytics.

> "In order to execute on our data analytics strategy, new technologies were needed in order to improve data engineering and boost data science productivity."
>
> **DANIEL SANCHEZ**
> Enterprise Data Architect at Banco Hipotecario

How Databricks Enables Data-Driven Innovation in Financial Services

Databricks provides financial services companies with a fully managed, cloud platform that accelerates innovation by unifying data engineering, data analysts and data science with the rest of the business. Databricks accelerates innovation by making all data actionable and unlocking new ways to explore and analyze data to drive new use cases that minimize operational risk, create a more engaging customer experience, and ensure higher return on equity.



databricks

How Databricks Enables Data-Driven Innovation in Financial Services

From BI to AI: Managing your entire data journey

Data analytics starts by aggregating and processing large collections of diverse data from various data sources. For many organizations, this requires a large infrastructure build-out. Databricks provides a reliable and scalable platform for your data pipelines, data lakes and data platforms. This helps you manage your full data journey, so you can ingest, process, store and expose data throughout your organization through machine learning or data analytics.

> “Databricks has provided one platform for our data teams to access and share data across ABN AMRO and deliver ML-based solutions that drive automation and insight throughout the company.”
>
> **STEFAN GROOT**
> Engineering Lead Data Analytics at ABN AMRO

DATA INGEST	DATA PIPELINES	DATA LAKES	DATA CONSUMERS
Pull data across all your different data sources, data storage locations and data types, including batch and streaming. Leverage a library of connectors, integrations and APIs for all your needs.	Run scalable and reliable data pipelines. Use Scala, Python, R or SQL to run processing jobs quickly on distributed Apache Spark™ runtimes, without having to worry about the undelying compute.	Build reliable data lakes at scale. Improve data quality, optimize storage performance and manage stored data, all while maintaining data lake compliance and security.	Use your data lake as a shared source of truth across data science, machine learning and business analytics teams — BI dashboards, production models and everything in-between.

Databricks empowers FSIs to unify all forms of data for downstream analytics and machine learning

databricks

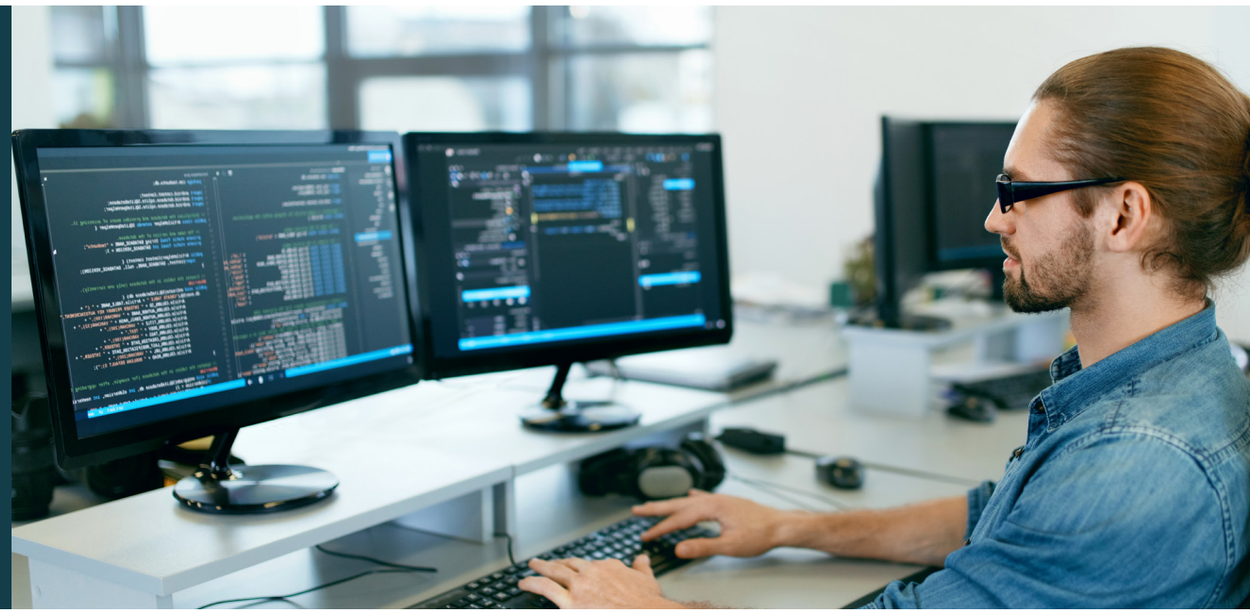
How Databricks Enables Data-Driven Innovation in Financial Services

Building performant and reliable data pipelines at scale

One of the most common problems organizations face when dealing with massive volumes of real-world data across disparate sources is that it can become unreliable, low quality and challenging to manage. Many organizations turn to data lakes to aggregate their big data cost-effectively, but this poses its own challenges.

Delta Lake is an open source technology that adds reliability and performance to your data lake and is natively integrated in Databricks. As part of The Linux Foundation, it has become one of the fastest growing, big data open-source projects. Delta Lake is a software layer that sits on top of your data lake, which enforces data quality to the data that enters your data lake. FSIs can land structured, semi-structured and unstructured data, both in batch and in real time, to a single Delta Lake to ensure that the supply of data is clean and usable. More importantly for FSIs, Delta Lake has the ability to go back in time and show a snapshot of data at a particular point in time. This feature unlocks many regulatory use cases from GDPR to CECL.

> “With Databricks, we are able to train models against all our data more quickly, resulting in more accurate pricing predictions that have had a material impact on revenue.”
>
> **BRYN CLARKE**
> Data Scientist at Nationwide



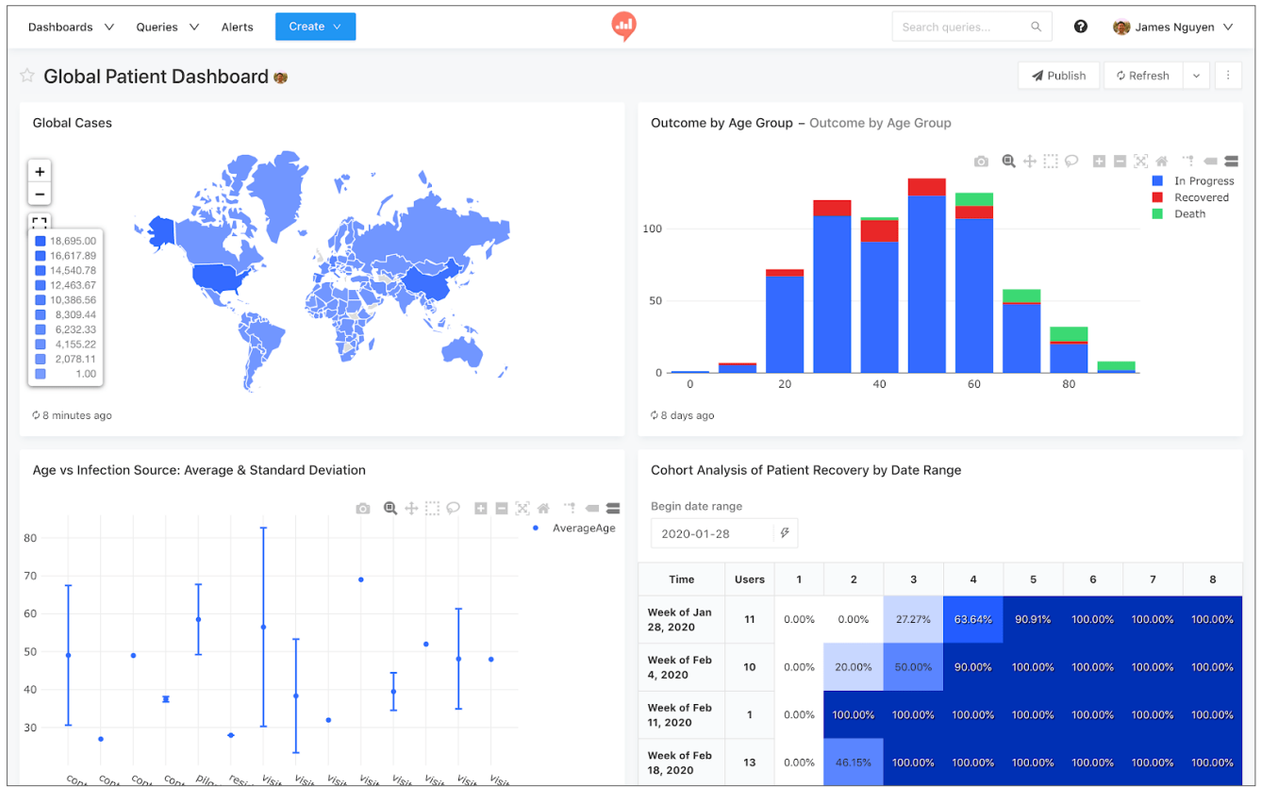
databricks

How Databricks Enables Data-Driven Innovation in Financial Services

Extracting business insights for analytics

With data centralized for easy access, Databricks enables you and your team of data analysts to easily and directly connect and query your most complete and recent data in the data lake with Delta Lake and Spark SQL. Connectors with popular BI tools like Tableau and Power BI allow your analysts to use their preferred BI visualization and reporting tools for real-time insights.

For a completely seamless experience, you can leverage Redash, a natively integrated visualization tool, to easily visualize and share your data via intuitive dashboards and queries.



How Databricks Enables Data-Driven Innovation in Financial Services

Bring data science and analytics teams together to accelerate innovation

Key to ensuring a rapid pace of data-driven innovation is to foster a collaborative environment that empowers data teams to work better together across the enterprise. When data teams work together effectively, they are able to more easily focus their ideas, skills and energy toward accomplishing amazing things. Through an interactive workspace, data engineers and scientists can easily collaborate on data, share models and code, and manage the entire machine learning lifecycle in one place. Databricks notebooks natively support Python, R, SQL and Scala so practitioners can work together with the languages and libraries of their choice and then push results to business stakeholders with built-in dashboards and visualizations.

Streamlining the machine learning lifecycle to create customer value

Successfully building and deploying a machine learning model can be difficult but essential to FSIs. MLflow is an open source framework that streamlines the machine learning lifecycle — allowing data scientists to reproduce a pipeline, compare the results of different versions, track what's running where, and redeploy and roll back updated models. MLflow is natively integrated into the Databricks Unified Data Analytics Platform, allowing your teams to seamlessly connect their data pipelines to their models in development and track them across the entire ML lifecycle.

All this is made possible by a Unified Data Analytics Platform that enables data scientists, bankers, traders, underwriters, risk managers and the like to explore big data and draw new insights.

> "Our ability to embed ML and AI in all aspects of our business has been crucial in creating more value for our clients. Azure Databricks and MLflow are core to our ability to deliver on this value."
>
> **ANURAG SEHGAL**
> Managing Director, Head of Global Markets at Credit Suisse

databricks

Creating Value for the Business With Data

Once data is reliable and performant, the next hurdle FSIs must overcome is the cumbersome *model delivery lifecycle* that is unique to the industry given its highly regulated nature. Simply put, the process of creating, auditing and deploying models in financial services can routinely take more than 12 months. The key is to unify both the data science and engineering challenges, enable collaboration and peer review, and bring transparency and reliability to both your model and data (through MLflow and Delta Lake). This ability to quickly enable commercial, responsible and explainable use of data — while adhering to existing risk frameworks, bank policies and regulations — is the key to driving value creation while safeguarding one of the most important assets: customer trust.

> “We’ve seen major improvements in the speed we have data available for analysis. We have a number of jobs that used to take 6 hours and now take only 6 seconds.”
>
> **ALESSIO BASSO**
> Chief Architect at HSBC

The Databricks Impact

Databricks helps companies automate infrastructure management, increase ETL performance at scale, and accelerate machine learning and analytics initiatives.

12x faster ETL pipelines

Impact: Faster time-to-market of new analytics insights and models.

+25% gains in productivity

Impact: More productive data scientists results in more AI innovation.

+47% Overall cost savings

Impact: Lower infrastructure costs boosts operational margins

databricks

Solution Accelerators

Based on best practices from our work with the leading brands, we've developed solution accelerators for common data analytics and machine learning use cases to save weeks or months of development time for your data engineers and data scientists.

MODERNIZE RISK MANAGEMENT

Traditional banks relying on on-premises infrastructure can no longer effectively manage risk. Moving to a cloud-based data analytics platform helps FSIs adopt a more agile approach to risk management by unifying data and AI.

DATA-DRIVEN ESG

ESG is widely considered a top-level initiative to achieve resilience and sustainable profitability in a rapidly evolving economic and environmental landscape. This solution provides companies and investors a holistic and data-driven view into ESG performance.

FRAUD PREVENTION WITH PREDICTIVE ANALYTICS

Curbing fraudulent or malicious behavior — from fraudulent securities trading to money laundering — is key to mitigating negative revenue impact. This solution provides ways to leverage data and machine learning to detect anomalies.

ACCESS OUR ACCELERATORS TO GET STARTED

databricks

CUSTOMER SPOTLIGHT

HSBC: Reinventing mobile banking with ML

Bank grade security and hyper-fast payment powered by data and ML HSBC

INDUSTRY:
Financial services

SOLUTION:
Customer 360, fraud detection, personalized experience, recommendation engine

TECHNICAL USE CASE:
Data ingest and ETL, machine learning, deep learning

As one of the largest international banks, HSBC is ushering in a new way to manage digital payments across mobile devices. They developed PayMe, a social app that facilitates cashless transactions between consumers and their networks instantly and securely. With over 39 million customers, HSBC struggled to overcome scalability limitations that blocked them from making data-driven decisions. With Databricks, they are able to scale data analytics and machine learning to feed customer-centric use cases including personalization, recommendations, network science, and fraud detection.

Data science and engineering struggled to leverage data

HSBC understands the massive opportunity for them to better serve their 39+ million customers through data and analytics. Seeing an opportunity to reinvent mobile payments, they developed the PayMe, a social payments app. Since its launch in their home market of Hong Kong, they have become the #1 app in the region amassing 1.8+ million users.

In an effort to provide their fast growing customer base the best possible mobile payments experience, they looked to data and machine learning to enable various desired use cases such as detecting fraudulent activity, customer 360 to inform marketing decisions, personalization, and more. However, building models that could deliver on these use cases in a secure, fast and scalable manner was easier said than done.

- **Slow data pipelines resulted in old data:** Legacy systems hampered their ability to process and analyze data at scale. They were required to manually export and sample data, which was time consuming. This resulted in the data being weeks old upon delivery to the data science team which blocked their ability to be predictive.
- **Manual data exporting and masking:** Legacy processes required a manual approval form to be filled out for every data request which was error-prone. Furthermore, the manual masking process was time consuming and did not adhere to strict data quality and protection rules.
- **Inefficient data science:** Data scientists worked in silos on their own machines and custom environments, limiting their ability to explore raw data and train models at scale. As a result, collaboration was poor and iteration on models were very slow.
- **Data analysts struggled to leverage data:** Needing access to subsets of structured data for business intelligence and reporting.

databricks

CUSTOMER SPOTLIGHT

HSBC: Reinventing mobile banking with ML

Bank grade security and hyper-fast payment powered by data and ML HSBC

Faster and more secure analytics and ML at scale

Through the use of NLP and machine learning, HSBC is able to quickly understand the intent behind each transaction within their PayMe app. This wide range of information is then used to inform various use cases from recommendations to customers to reducing anomalous activity.

With Azure Databricks, they are able to unify data analytics across data engineering, data science, and analysts.

- **Improved operational efficiency:** features such as auto-scaling clusters and support for Delta Lake has improved operations from data ingest to managing the entire machine learning lifecycle.
- **Real time data masking with delta lake:** With Databricks and Delta Lake, HSBC was able to securely provide anonymized production data in real-time to data science and data analyst teams.
- **Performant and scalable data pipelines with Delta Lake:** This has enabled them to perform real-time data processing for downstream analytics and machine learning.
- **Collaboration across data science and engineering:** Enables faster data discovery, iterative feature engineering, and rapid model development and training.

Richer insights leads to the #1 app

Databricks provides HSBC with a unified data analytics platform that centralizes all aspects of their analytics process from data engineering to the productionization of ML models that deliver richer business insights.

- **Faster data pipelines:** Automating processes and increased data processing from 6 hours to 6 seconds for complex analytics.
- **Descriptive to predictive:** Ability to train models against their entire dataset, has empowered them to deploy predictive models to feed various use cases.
- **From 14 databases to 1 Delta Lake:** Moved from 14 read replica databases to a single unified data store with Delta Lake.
- **PayMe is #1 app in Hong Kong:** 60% market share of the Hong Kong market making PayMe the #1 app.
- **Improved consumer engagement:** Ability to leverage network science to understand customer connections has resulted in a 4.5x improvement in engagement levels with the PayMe app.

170+	6	1	4.5X
PBs of data in data centers across 21 countries	Seconds to perform complex analytics compared to 6 hours	Delta Lake has replaced 14 databases	Improvement in engagement on the app

databricks

Conclusion

Data is at the core of nearly every innovation in the financial services industry. Databricks enables FSIs across banking, insurance, capital markets and wealth management to harness the power of data and analytics to solve strategic challenges and make smarter decisions that minimize risk, prevent fraudulent behavior and drive sustainable value creation.

Get started with a free trial of Databricks and start building data applications today

START YOUR FREE TRIAL

Contact us for a personalized demo **databricks.com/contract**

databricks