Big Book of Data and AI
Use Cases for the Public Sector

Best practices, customer stories and solution templates for government agencies interested in building on the Lakehouse

databricks

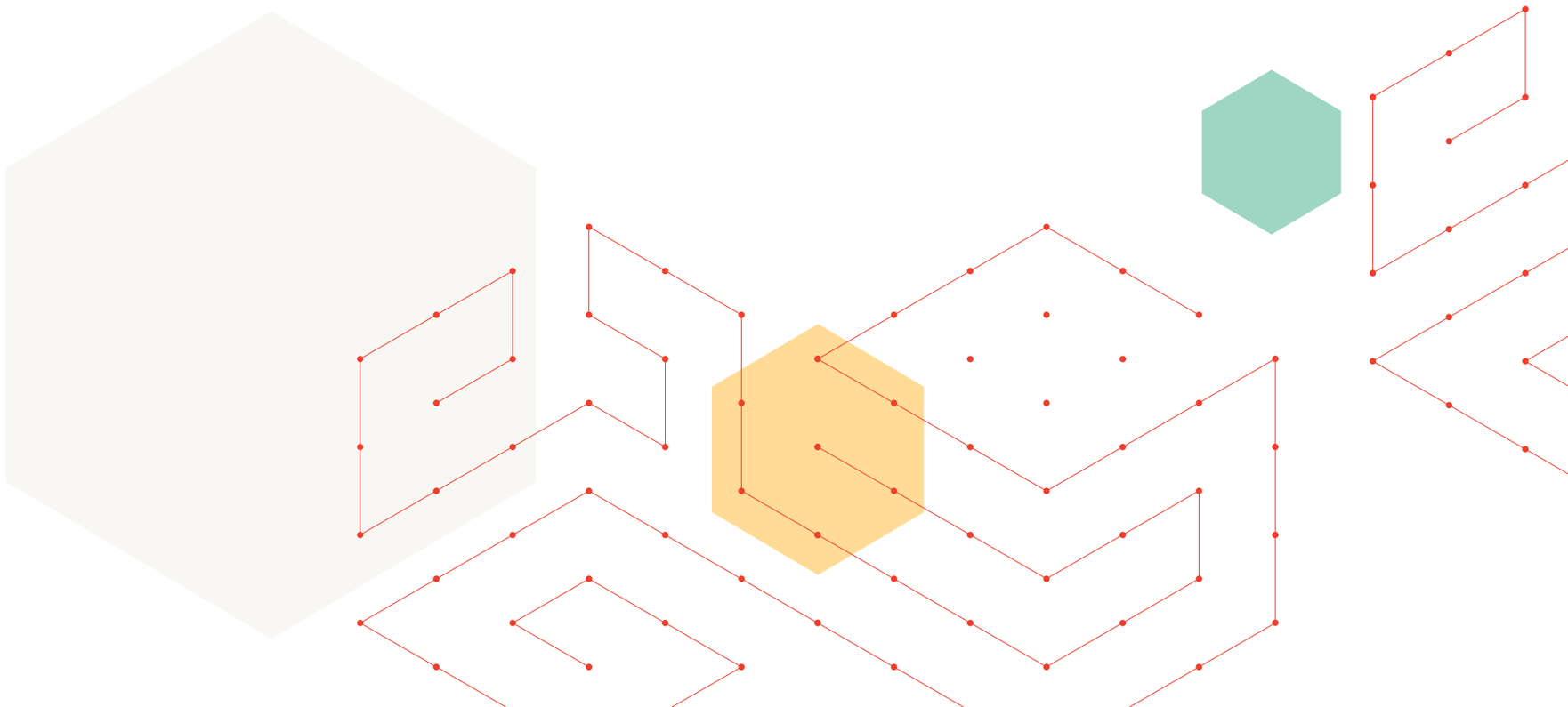
Contents

The State of Data and AI in the Government

Over the last decade, data and AI have redefined every industry on the planet. Retailers have improved the shopping experience with personalized recommendations, financial institutions have strengthened risk management through the use of advanced analytics, and the healthcare industry is tapping into the power of machine learning to predict and prevent chronic disease. The public sector is no exception.

In 2018, the U.S. Federal Government embarked on one of its most ambitious efforts since putting a man on the moon — embedding data into all aspects of decision-making. By enacting the Evidence-Based Policymaking Act of 2018, Congress set in motion requirements for agencies to modernize their data and analytics capabilities, including the appointment of agency-level chief data officers. A year later came the Federal Data Strategy, which provided further guidance for how agencies should manage and use data by 2030.

With all of this guidance, agencies are starting to make meaningful improvements to their data strategy, but when it comes to innovating with data, agencies still lag behind the private sector. This begs the question: what's standing in the way? The hurdles aren't due to a lack of effort on the part of agency leaders. In fact, they can largely be attributed to a patchwork of legacy technologies that have been amassed over the last 30 to 40 years. While these hurdles stand in the way, a number of innovative agencies are making significant progress as they embrace new data and AI capabilities.



databricks

Federal spending on artificial intelligence rose to nearly $1 billion in 2020, up 50% from 2018. There's a good reason for this level of spend: Deloitte recently published a report, "AI-augmented Government," that estimates the federal government could free up as many as 1.2 billion hours of work and save up to $41.1 billion annually through the use of AI-driven automation. Early adopters of advanced analytics are starting to see the fruits of their labor. For example, USCIS modernized their analytics stack on Databricks to accelerate insights on applicants by 24x, automate the processing of millions of applications, and reduce appointment no-show rates with predictive analytics. The Orange County Courts also recently shared how they are automating legacy paper-based workflows with machine learning.

An increased focus on cloud, analytics and AI = operational efficiency

$1B	1. AI/ML 2. Data Analytics 3. Cloud	**$41B+**
Data and AI Research and Development Initiative	**TOP PRIORITIES** Government CIOs' top game-changing technologies	Estimated government savings from data-driven automation
U.S. Government	Gartner	Deloitte.

In this eBook, we explore the hurdles of legacy technologies and how a modern data lakehouse can help agencies unlock innovative data and analytics use cases at all levels of government. Over the following seven example use cases, covering everything from cyber threat detection to improving public health, we demonstrate how the Databricks Lakehouse for Public Sector is critical to improving citizen services and delivering on mission objectives. This guide also includes resources in the form of Solution Accelerators, reference architectures and real-world customer stories to help as you embark on your own journey to drive a safer and more prosperous nation through the use of data and AI.

The Need for a Modern Data Architecture

Government agencies are now turning to the cloud and modern data technologies to federate and make sense of their massive volumes of data. Building on that foundation, agencies are starting to adopt advanced analytics and AI to automate costly, outdated and resource-intensive operations as well as improve decision-making with predictive insights that can better keep pace with the dynamic needs of citizens and global communities. That being said, there are a number of barriers standing in their way.

Common challenges

Many government agencies are burdened with a legacy IT infrastructure that is built with on-premises data warehouses that are complex to maintain, are costly to scale as compute is coupled with storage, and lack support for unstructured data and advanced analytics. This severely inhibits data-driven innovation. Maintaining these systems requires a massive investment of both time and money compared to modern cloud-based systems and creates a number of avoidable challenges:

Lack of reliability

Siloed systems result in data replication as teams spin up new data marts to support their one-off use cases. Without a single source of truth, teams struggle with data inconsistencies, which can result in inaccurate analysis and model performance that is only compounded over time.

Lack of agility

Disjointed analytics tools and legacy infrastructure hinder the ability of teams to conduct real-time analytics. Most data processing in the government is often done in weekly or daily batches, but decision-making needs to happen in real time. Critical events like cyber attacks and health pandemics can't wait a week.

Lack of citizen insights

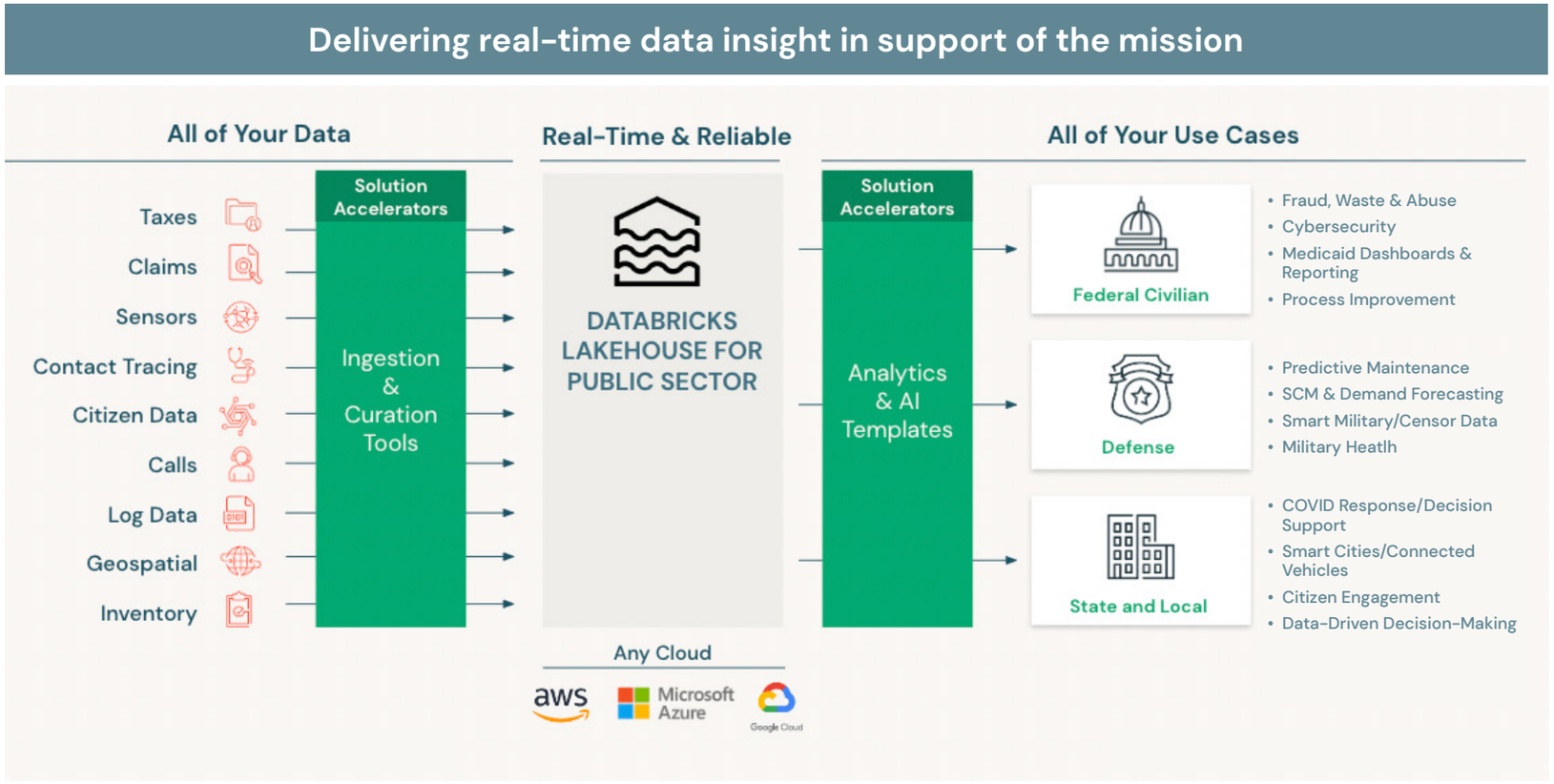
When data is siloed, teams get an incomplete view of the citizen, resulting in missed opportunities to improve the delivery of services that impact the quality of life for their constituents.

Lack of productivity

Data scientists and data analysts alike must have the right tool set to collaboratively investigate, extract and report meaningful insights from their data. Unfortunately, data silos lead to organizational silos, which make collaboration inside an agency as well as between agencies very difficult. With different groups of data teams leveraging their own coding and analytical tools, communicating insights and working across teams — let alone across agencies — is almost impossible. This lack of collaboration can drastically limit the capabilities of any data analytics or AI initiative.

Introducing the Lakehouse for Public Sector

The reason that the Databricks Lakehouse is able to deliver the simplicity, flexibility and speed that a government agency requires is that it fundamentally reimagines the modern data architecture. Databricks provides federal, state and local agencies with a cloud-native Lakehouse Platform that combines the best of data warehouses and data lakes — to store and manage all your data for all your analytics workloads. With this modern architecture, agencies can federate all their data and democratize access for downstream use cases, empowering their teams to deliver on their mission objectives by unlocking the full potential of their data.



Federate all of your agency's data

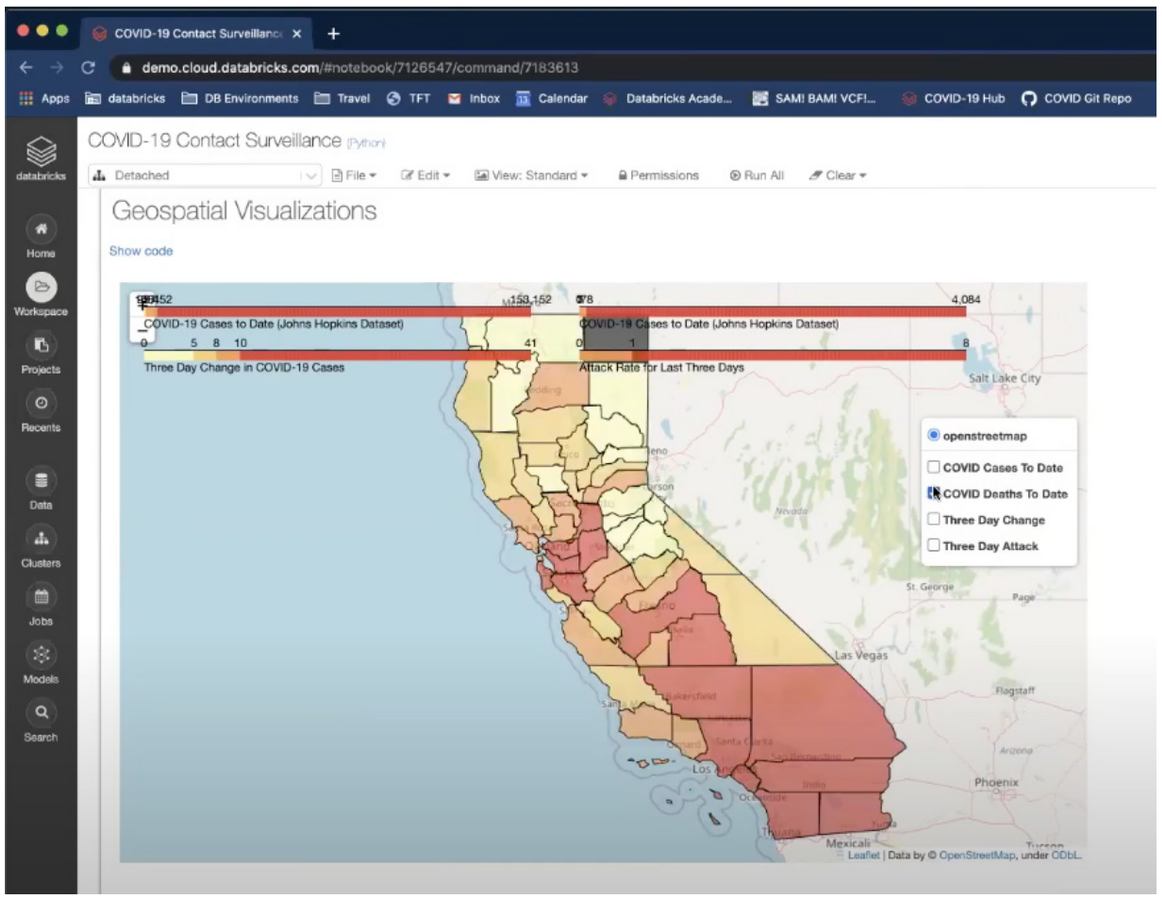
Any type of data can be stored because, like a data lake, the Databricks Lakehouse is built using the low-cost object storage supported by cloud providers. Leveraging this capability helps break down the data silos that hinder efforts to aggregate data for advanced analytics (e.g., predictive maintenance) or compute-intensive workloads like detecting cyber threats across billions of signals. Probably even more important is the ability of the lakehouse architecture to travel back in time, ensuring full audit compliance and high governance standards for analytics and AI.

Power real-time decision-making

Streaming use cases such as IoT analytics or disease spread tracking is simpler to support because the lakehouse uses Apache Spark™ as the data processing engine and Delta Lake as a storage layer. With Spark, you can toggle between batch and streaming workloads with just a line of code. With Delta Lake, native support for ACID transactions means that you can deploy streaming workloads without the overhead of common reliability and performance issues. These capabilities make real-time analytics possible.

Unlock collaborative analytics for all personas

The Databricks Lakehouse for Public Sector is your one-stop shop for all your analytics and AI. The platform includes a business intelligence capability — Databricks SQL — that empowers data analysts to query and run reports against all of an agency's unified data. Databricks SQL integrates with BI tools like Tableau and Microsoft Power BI and complements any existing BI tools with a SQL-native interface, allowing data analysts and data scientists to query data directly within Databricks and build powerful dashboards.



databricks

Deliver on your mission with predictive insights

In the same environment, data scientists can build, share and collaborate on machine learning models for advanced use cases like fraud detection or geospatial analytics. Additionally, MLflow, an open source toolkit for managing the ML lifecycle, is built into the Lakehouse so data scientists can manage everything in one place. Databricks natively supports Python, R, SQL and Scala so practitioners can work together with the languages and libraries of their choice, reducing the need for separate tools. With these capabilities, data teams can turn insights from real-world data into powerful visualizations designed for machine learning. Visualizations can then be turned into interactive dashboards to share insights with peers across agencies, policymakers, regulators and decision-makers.

Customers That Innovate With Databricks Lakehouse for Public Sector

Some of the top government agencies in the world turn to the Databricks Lakehouse for Public Sector to bring analytics and AI-driven automation and innovation to the communities they serve.

USE CASE: Cybersecurity

Overview

Cyberattacks from bad actors and nation states are a huge and growing threat to government agencies. Recent large-scale attacks like the ones on SolarWinds, log4j, Colonial Pipeline and HAFNIUM highlight the sophistication and increasing frequency of broad-reaching cyberattacks. Data breaches cost the federal government more than $4 million per incident in 2021 and threaten national security. Staying ahead of the next threat requires continuous monitoring of security data from an agency's entire attack surface before, during and after an incident.

Challenges

Scaling existing SIEM solutions

Agencies looking to expand existing SIEM tools for today's petabytes of data can expect increased licensing, storage, compute and integration resources resulting in tens of millions of dollars in additional costs per year.

Rules-based systems

Many legacy SIEM tools lack the critical analytics capabilities — such as advanced analytics, graph processing and machine learning — needed to detect unknown threat patterns or deliver on a broader set of security use cases like behavioral analytics.

Limited window of data

Given the high cost of storage, most agencies retain only a few weeks of threat data. This can be a real problem in scenarios where a perpetrator gains access to a network but waits months before doing anything malicious. Without a long historical record, security teams can't analyze cyberattacks over long-term horizons or conduct deep forensic reviews.

Solution overview

For government agencies that are ready to modernize their security data infrastructure and analyze data at petabyte-scale more cost-effectively, Databricks provides an open lakehouse platform that augments existing SIEMs to help democratize access to data for downstream analytics and AI. Built on Apache Spark and Delta Lake, Databricks is optimized to process large volumes of streaming and historic data for real-time threat analysis and incident response. Security teams can query threat data going years into the past in just minutes and build ML models to detect new threat patterns and reduce false positives. Additionally, Databricks created a Splunk-certified add-on to augment Splunk for Enterprise Security (ES) for cost-efficient log and retention expansion.

How to get started

Solution Accelerator: Detect Criminal Threats Using DNS Analytics

Detecting criminals and nation states through DNS analytics. In order to address common cybersecurity challenges such as deployment complexity, tech limitation and cost, security teams need a real-time data analytics platform that can handle cloud scale, analyze data wherever it is, natively support streaming and batch analytics, and have collaborative content development capabilities.

Solution Accelerator: Databricks Add-On for Splunk

Designed for cloud-scale security operations, the add-on provides Splunk analysts with access to all data stored in the Lakehouse. Bidirectional pipelines between Splunk and Databricks allow agency analysts to integrate directly into Splunk visualizations and security workflows.

Customer story

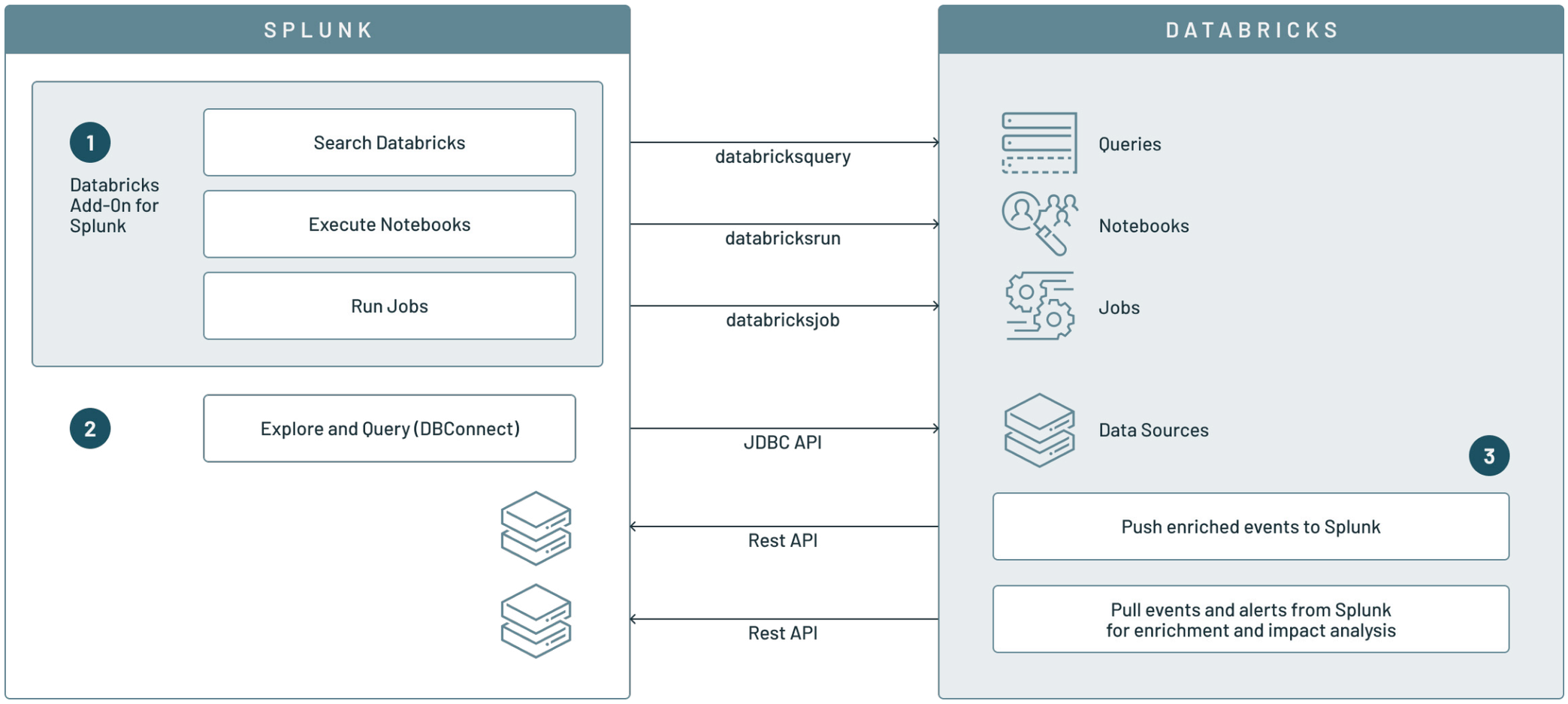
HSBC

WATCH THE VIDEO >

Fighting Cyber Threats in Real Time

Since partnering with Databricks, HSBC has reduced costs, accelerated threat detection and response, and improved their security posture. Not only can they process all of their required data, but they've also increased online query retention from just days to months at petabyte scale. HSBC is now able to execute 2-3x more threat hunts per analyst.

Reference architecture



USE CASE: Predictive Maintenance

Overview

Predictive maintenance is oftentimes associated with the manufacturing sector, but in reality it extends far beyond the factory floor. Consider this for a moment: the U.S. Government operates a fleet of over 640,000 vehicles including public buses, postal delivery trucks, drones, helicopters and jet fighters. Many of these vehicles — like multimillion-dollar aircraft — contain sensors that generate massive amounts of data on the use and conditions of various components. And it's not just vehicles. Modern public utilities stream data through connected IoT devices. All of this data can be analyzed to identify the root cause of a failure and predict future maintenance, helping to avoid costly repairs and critical assets from being out of service.

Challenges

Managing IoT data at scale

With billions of sensors generating information, most data systems are unable to handle the sheer volume of data. Before agencies can even start analyzing their data, legacy data warehouse–based tools require preprocessing of data, making real-time analysis impossible.

Integrating unstructured data

Equipment data doesn't just come in the form of IoT data. Agencies can gather rich unstructured signals like audio, visual (e.g., video inspections) and text (e.g., maintenance logs). Most legacy data architectures are unable to integrate structured and unstructured data sources.

Operationalizing machine learning

Most agencies lack the advanced analytics tools needed to build models that can predict potential equipment failures. Those that do typically have their data scientists working in a siloed set of tools, resulting in unnecessary data replication and inefficient workflows.

Solution overview

The Databricks Lakehouse is tailor-made for building IoT applications at scale. With Databricks, agencies can easily manage large streaming volumes of small files, with ACID transaction guarantees and reduced job fails compared to traditional data warehouse architectures. Additionally, the Lakehouse is cloud native and built on Apache Spark, so scaling for petabytes of data is not an issue. With the Lakehouse, agencies can bring together all of their structured and unstructured data with a unified set of tooling for data engineering, model building and production rollout. With these capabilities, operations teams can quickly detect and act on pending equipment failures before they affect performance.

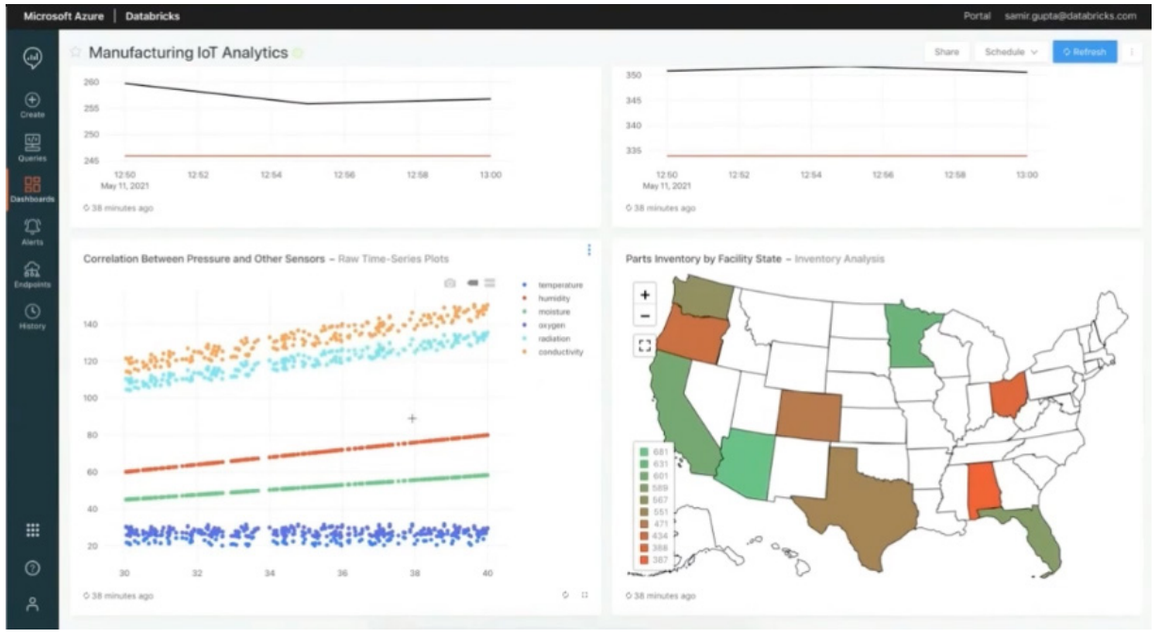
How to get started

Solution Accelerator: Predictive Maintenance

Learn how to ingest real-time IoT data from field devices, perform complex time series processing in Delta Lake and leverage machine learning to build predictive maintenance models.

- Part 1: Use case overview
- Part 2: Ingest real-time IoT data and perform time series processing
- Part 3: Using ML to predict maintenance.

Watch the Demo:
Predictive Maintenance on Azure Databricks



Customer story

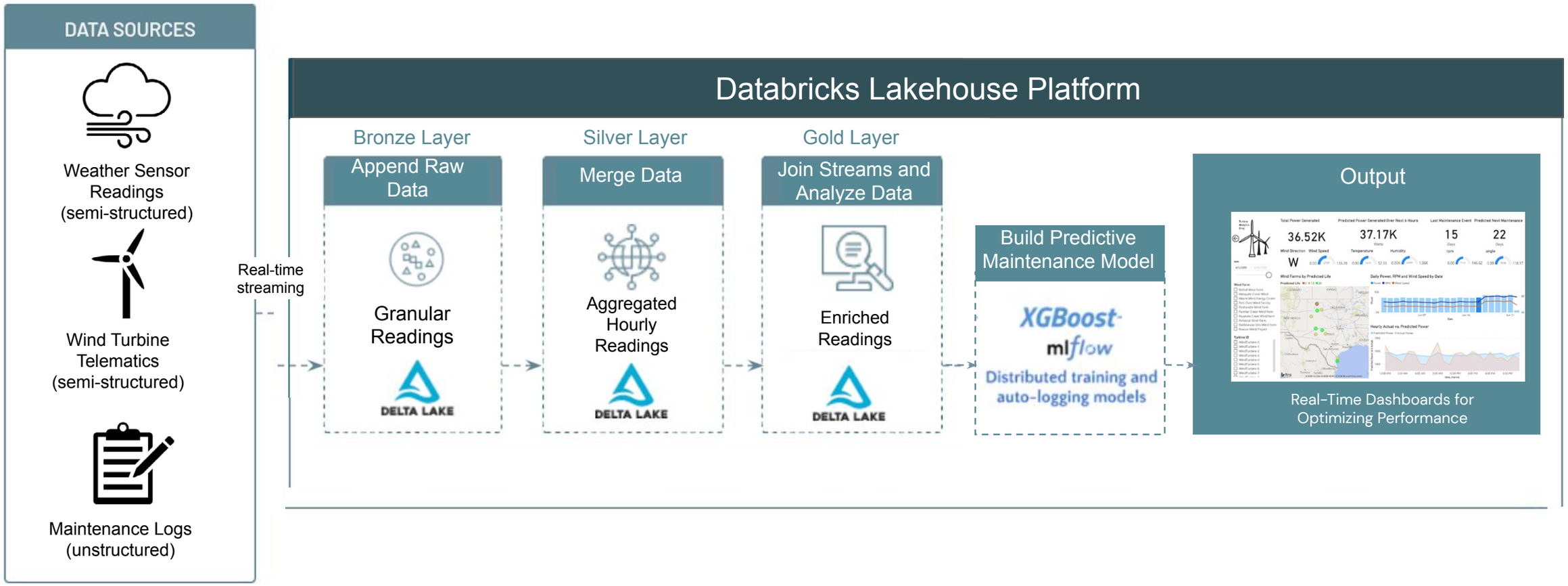
LEARN MORE >

Protecting the Water Supply for 700,000 Residents

Utilizing machine learning for predictive analytics to help stop water main breaks before they occur, potentially saving hundreds of thousands of dollars in repairs while reducing service interruption.

databricks

Reference architecture



USE CASE: Fraud Detection

Overview

According to McKinsey & Company, more than half of the federal government's monetary losses to fraud, waste and abuse go undetected and total tens of billions of dollars. Financial fraud comes in many forms, from individuals taking advantage of relief programs to complex networks of criminal organizations working together to falsify medical claims and rebate forms. Investigative teams hoping to stay ahead of fraudsters need advanced analytics techniques so they can detect anomalous behavior buried in a sea of data.

Challenges

Lack of machine learning

A rules-based approach is not enough. Bad actors are getting more and more sophisticated in how they take advantage of government programs, necessitating an AI-driven approach.

Unreliable data

Getting high-quality, clean data and maintaining a rich feature store is critical for identifying ever-evolving fraud patterns while maintaining a strict record of previous data points.

Analytics at scale

Training complex ML models with hundreds of features on gigabytes of structured, semi-structured and unstructured data can be impossible without a highly scalable and distributed infrastructure.

Solution overview

The Databricks Lakehouse enables teams to develop complex ML models with high governance standards and bridge the gap between data science and technology to address the challenge of analyzing large volumes of data at scale — 40 billion financial transactions a year are made in the United States alone. Additionally, Databricks makes it possible to combine modern AI techniques with the legacy rules-based methods that underpin current approaches to fraud detection all within a common and efficient Spark-based orchestration engine.

How to get started

Solution Accelerator: Fraud Detection

Due to an ever-changing landscape, building a financial fraud detection framework often goes beyond just creating a highly accurate machine learning model. Oftentimes it involves a complex-decision science setup that combines a rules engine with a need for a robust and scalable machine learning platform. In this example, we show how to build a holistic fraud detection solution on Databricks using data from a financial institution.

Customer story

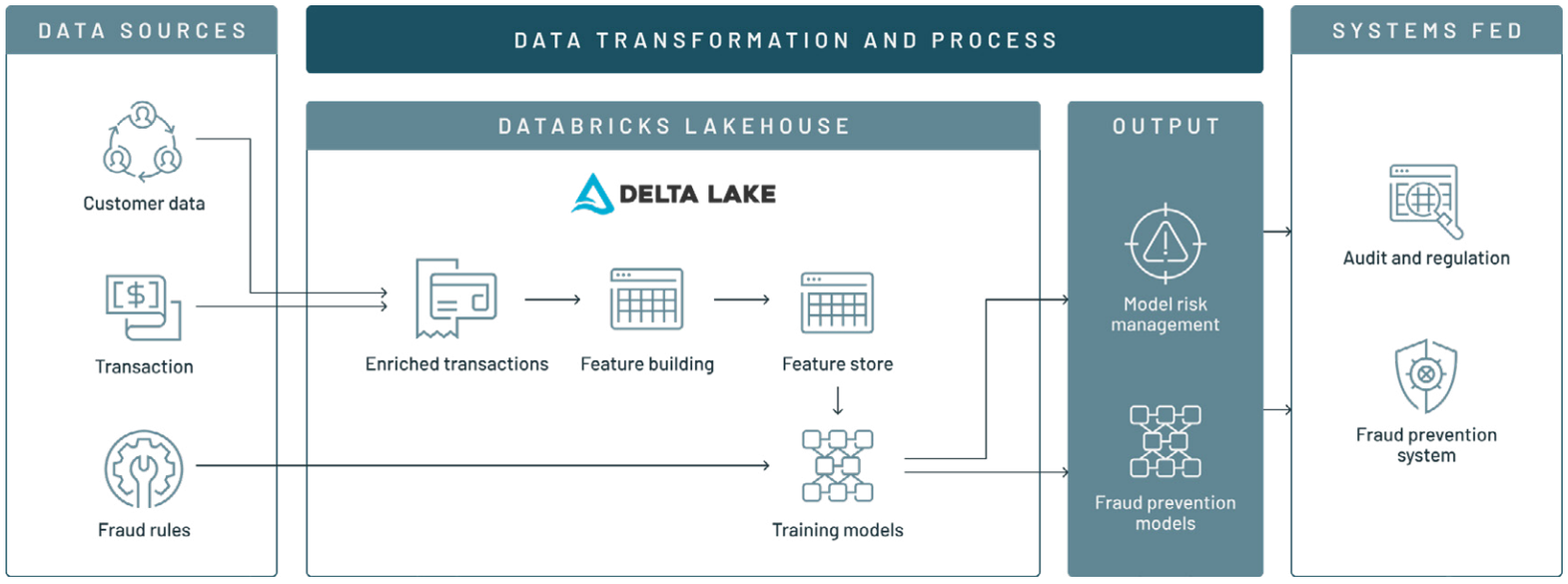
FINRA®

WATCH THE VIDEO >

Identifying Financial Fraud at Scale

Processes hundreds of billions of market events per day on the Databricks Lakehouse and uses the power of machine learning to identify illicit activity in near real-time.

Reference architecture



USE CASE: Money Laundering

Overview

Approximately $300 billion is laundered through the United States each year, and with criminal organizations — both at home and abroad — implementing increasingly sophisticated methods for laundering funds, it's getting harder to stop. While the federal government continues to apply pressure on the financial sector through heightened regulation, more is needed to combat laundering. Modern AI techniques such as graph analytics and computer vision can be used to process different types of structured (e.g., financial transactions) and unstructured (e.g., real estate images) data and identify illicit behavior. This allows investigative teams to automate labor-intensive activities like confirming a residential address or reviewing transaction histories, and instead dig into priority threats.

Challenges

Complex data science

Modern anti-money laundering (AML) practices require multiple ML capabilities such as entity resolution, computer vision and graph analytics on entity metadata, which is typically not supported by any one data platform.

Model transparency

Although AI can be used to address many money laundering use cases, the lack of transparency in the development of ML models offers little explainability, inhibiting broader adoption.

Time-consuming false positives

Any reported suspicious activity must be investigated manually to ensure accuracy. Many legacy solutions generate a high number of false positives or fail to identify unknown patterns, resulting in wasted effort by investigators.

Solution overview

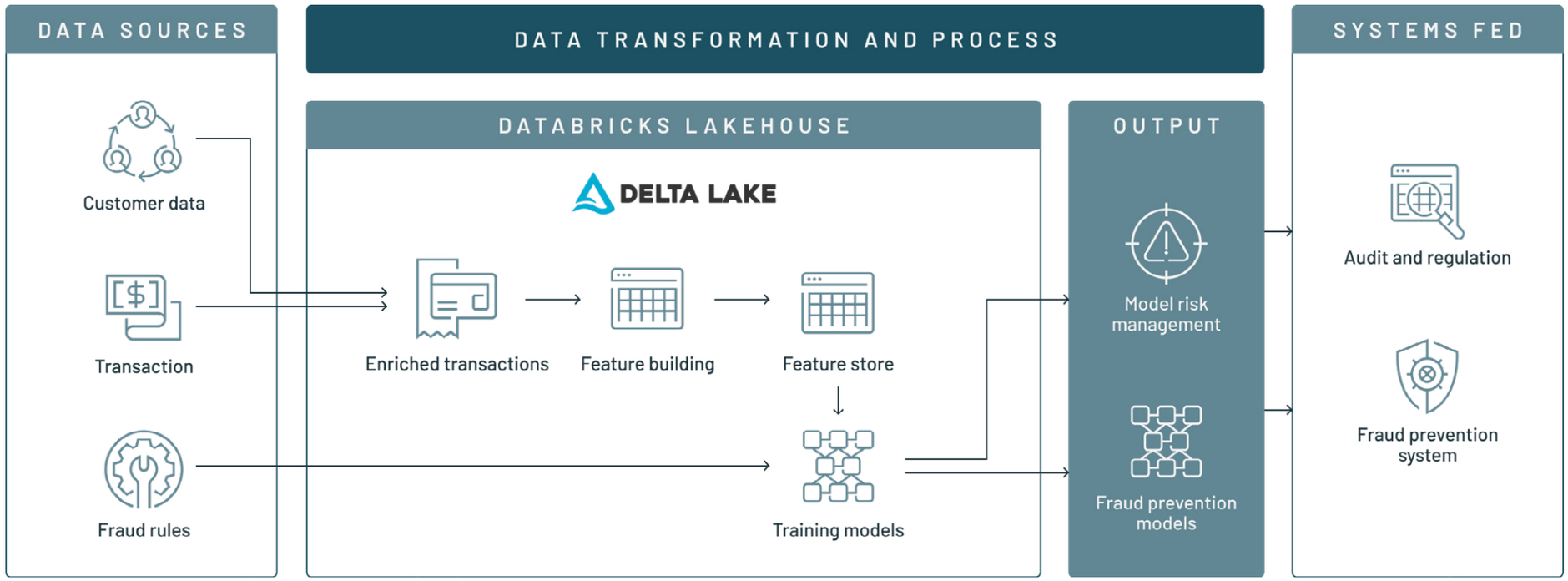
AML solutions face the operational burden of processing billions of transactions a day. The Databricks Lakehouse Platform combines the low storage cost benefits of cloud data lakes with the robust transaction capabilities of data warehouses, making it the ideal foundation for building AML analytics at massive scale. At the core of Databricks is Delta Lake, which can store and combine both unstructured and structured data to build entity relationships; moreover, Databricks Delta Engine provides efficient access using the new Photon compute to speed up BI queries on tables spanning billions of transactions. On top of these capabilities, ML is a first-class citizen in the Lakehouse, which means analysts and data scientists do not waste time subsampling or moving data to share dashboards and stay one step ahead of bad actors.

How to get started

Solution Accelerator: Modern Anti-Money Laundering Techniques

Current anti-money laundering practices bear little resemblance to those of the last decade. In today's digital world, financial institutions are processing billions of transactions daily, increasing the surface area of money laundering. With this accelerator, we demonstrate how to build a scalable AML solution on the Lakehouse Platform leveraging a series of next-gen machine learning techniques including NLP, computer vision, entity resolution and graph analytics. This approach helps teams better adapt to the reality of modern laundering practices.

Reference architecture



USE CASE: Entity Analytics

Overview

Entity analytics aims to connect disparate data sources to build a full view of a person or an organization. This has many applications in the public sector, such as fraud detection, national security and population health. For example, Medicare fraud teams need to understand which prescriptions are filled, claims filed and facilities visited across geographies to uncover suspicious behavior. Before teams can even look for suspicious behavior, they must first determine which records are associated. In the United States, nearly 50,000 people share the name John Smith (and there are thousands of others with similar names). Imagine trying to identify the right John Smith for this type of analysis. That's no easy task.

Challenges

Disjointed data

Managing complex and brittle ETL pipelines in order to cleanse and join data across siloed systems and data stores.

Compute intensive

Identifying related entities across population-level data sets requires massive compute power that far outstrips legacy on-prem data architectures.

No machine learning capabilities

Entity resolution typically relies on basic rules-based logic to compare records (e.g., matching on name and address), but with messy, large volumes of data, advanced analytics is needed to improve accuracy and accelerate efforts.

Solution overview

The Databricks Lakehouse is an ideal platform for building entity analytics at scale. With support for a wide range of data formats and a rich and extensible set of data transformation and ML capabilities, Databricks enables agencies to bring together all of their data in a central location and move beyond simple rules-based methods for entity resolution. Data teams can easily explore different machine learning techniques like natural language processing, classification and graph analytics to automate entity matching. And one-click provisioning and deprovisioning of cloud resources makes it easy for teams to cost-effectively allocate the necessary compute resources for any size job so they can uncover findings faster.

How to get started

Virtual Workshop: Entity Analytics

Learn from Databricks experts on how entity analytics is being deployed in the public sector and watch a demo that shows how to use ML to link payments and treatments across millions of records in a public CMS data set.

Solution Accelerator:
Machine Learning-Based Item Matching

While focused on retail, this accelerator has applications for any organization working on entity matching, especially as it relates to items that might be stored across locations. In this notebook, we demonstrate how to use machine learning and the Databricks Lakehouse Platform to resolve differences between product definitions and descriptions, and determine which items are likely pairs and which are distinct across disparate data sets.

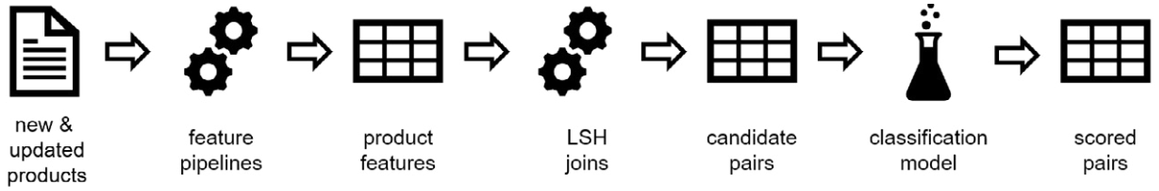
Customer story

CMS CENTERS FOR MEDICARE & MEDICAID SERVICES

WATCH THE VIDEO >

In this talk, NewWave shares the specifics on CMS's entity resolution use case, the ML necessary for this data and the unique uses of Databricks in providing this capability.

Sample workflow



USE CASE: Geospatial Analytics

Overview

Every day billions of handheld and IoT devices, along with thousands of airborne and satellite remote sensing platforms, generate hundreds of exabytes of location-aware data. This boom of geospatial big data combined with advancements in machine learning is enabling government agencies to develop new capabilities. The potential use cases for geospatial analytics and AI touch every part of the government, including disaster recovery (e.g., flood/earthquake mapping), defense and intel (e.g., detecting threats using drone footage), infrastructure (e.g., public transportation planning), civilian safety (e.g., crime prediction), public health (e.g., disease spread tracking), and much more. Every agency at the state and federal level needs to consider how they can tap into geospatial data.

Challenges

Massive volumes of geospatial data

With the proliferation of low-cost sensor arrays, GPS technologies and high-resolution imaging organizations are collecting tens of TBs of geospatial data daily, outpacing their ability to store and process this data at scale.

Compute-intensive spatial workloads

Geospatial data is complex in structure, with various formats not well suited for legacy data warehouses, as well as being compute intensive, with geospatial-specific transformations and queries requiring hours and hours of compute.

Broad range of analytics capabilities

Enterprises require a diverse set of data applications — including SQL-based analytics, real-time monitoring, data science and machine learning — to support geospatial workloads given the diverse nature of the data and use cases.

Solution overview

With Delta Lake at the core, the Databricks Lakehouse is ideal for geospatial workloads, as it provides a single source of truth for all types of structured, unstructured, streaming and batch data, enabling seamless spatio-temporal unification and cross-querying with tabular and raster-based data. Built on Apache Spark, the Lakehouse easily scales for data sets consisting of billions of rows of data with distributed processing in the cloud. To expand on the core capabilities of the Lakehouse, Databricks has introduced the Mosaic library, an extension to the Apache Spark framework, built for fast and easy processing of large geospatial data sets. Popular frameworks such as Apache Sedona or GeoMesa can still be used alongside Mosaic, and because Mosaic sits on top of Lakehouse architecture, it unlocks AI/ML and advanced analytics capabilities to support all types of geospatial use cases.

How to get started

Solution Accelerator:
Mosaic for Geospatial Analytics

Build a Lakehouse to support all of your geospatial analytics and AI use cases with the Mosaic library. Mosaic provides a number of capabilities including easy conversion between common spatial data encodings, constructors to easily generate new geometries from Spark native data types, many of the OGC SQL standard ST_ functions implemented as Spark Expressions for transforming, aggregating and joining spatial data sets, and optimizations for performing point-in-polygon joins using an approach we codeveloped with Ordnance Survey — all provided with the flexibility of a Scala, SQL or Python API.

Virtual Workshop: Geospatial Analytics and AI at Scale

Learn how to build powerful geospatial insights and visualizations with a Lakehouse for all your geospatial data processing, analytics and AI.

Customer story

U.S. Department of Transportation

WATCH THE VIDEO >

Analyzing Flight Data to Improve Aviation

To help airlines better serve their millions of passengers, USDOT built a modern analytics architecture on Databricks that incorporates data such as weather, flight, aeronautical and surveillance information. With this new platform, they reduced compute costs by 90% and can now power use cases such as predicting air cargo traffic patterns, flight delays and the financial impact of flight cancellations.

Customer story

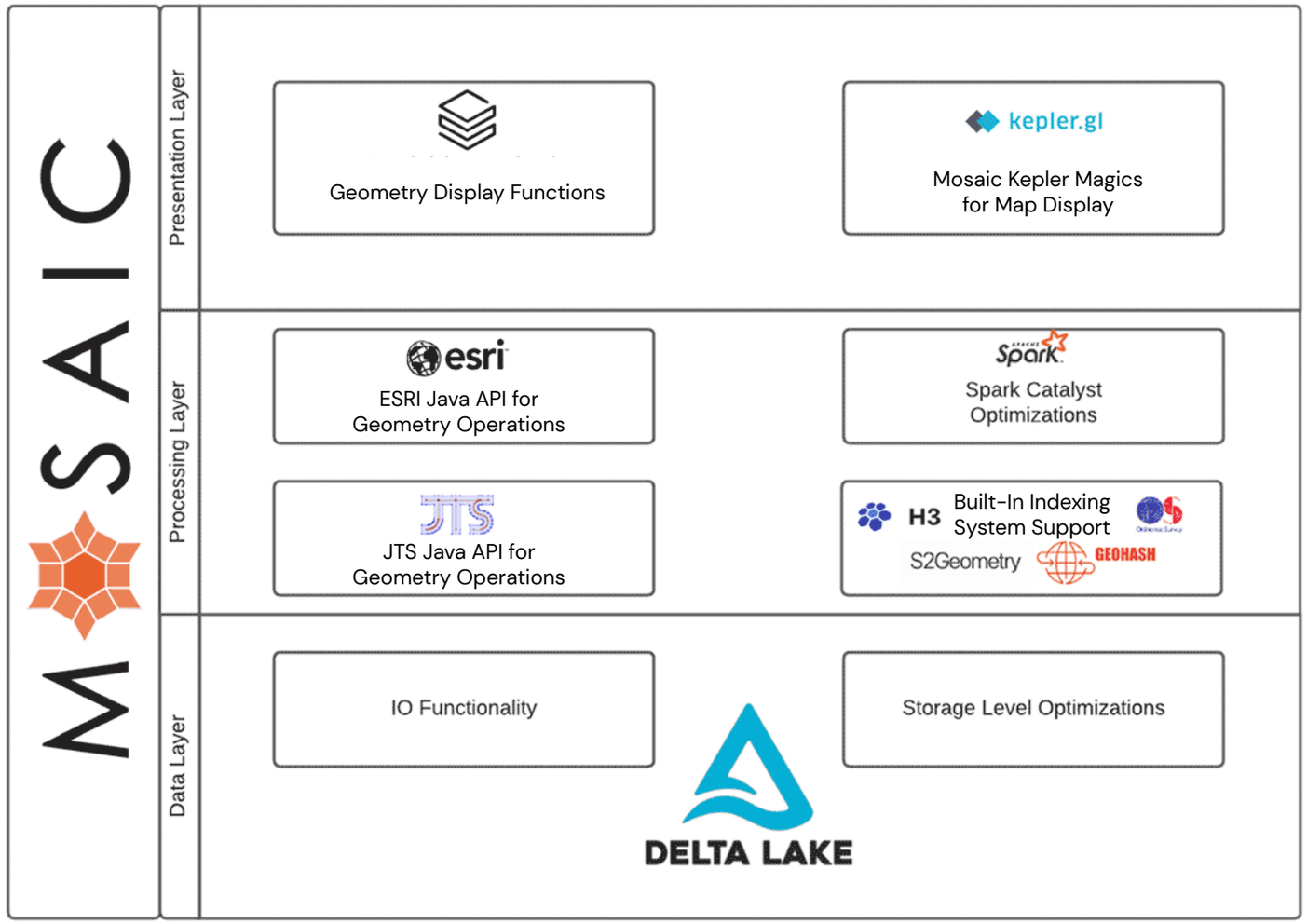
Stantec

WATCH THE VIDEO >

Customer Story: Flood Prediction With Machine Learning

In an effort to improve the safety of civil projects, Stantec built a machine learning model on Databricks leveraging large volumes of weather and geological data — oftentimes consisting of trillions of data points — to predict the impact of flash floods on various regions and adjust civil planning accordingly.

Reference architecture



USE CASE: Public Health Management

Overview

In their lifetime, every human is expected to generate a million gigabytes of health data spanning electronic health records, medical images, claims, wearable data, genomics and more. This data is critical to understanding the health of the individual, but when aggregated and analyzed across large populations, government agencies can glean important insights like disease trends, the impact of various treatment guidelines and the effectiveness of resources. By adding in Social Determinants of Health (SDOH) data — such as geographical location, income level, education, housing — agencies can better identify underserved communities and the critical factors that contribute to positive health outcomes.

Challenges

Rapidly growing health data

Healthcare data is growing exponentially. Unfortunately, legacy on-premises data architectures are complex to manage and too costly to scale for population-scale analytics.

Fragmented patient data

It is widely accepted that over 80% of medical data is unstructured, yet most organizations still focus their attention on data warehouses designed to only support structured data and SQL-based analytics.

Complexities of ML in healthcare

The legacy analytics platforms that underpin healthcare lack the robust data science capabilities needed for predictive health use cases like disease risk scoring. There's also the challenge of managing reproducibility, which is critical when building ML models that can impact patient outcomes.

Solution overview

The Databricks Lakehouse enables public health agencies to bring together all their research and patient data in a HIPAA-certified environment and marry it with powerful analytics and AI capabilities to deliver real-time and predictive insights at population scale. The Lakehouse eliminates the need for legacy data architectures, which have historically inhibited innovation in patient care by creating data silos and making advanced analytics difficult. Databricks led open source projects — like Glow for genomics and Smolder for EHR data — that make it easy to ingest and prepare healthcare-specific data modalities for downstream analytics.

How to get started

Solution Accelerator: NLP for Healthcare

Our joint solutions with John Snow Labs bring together the power of Spark NLP for Healthcare with the collaborative analytics and AI capabilities of Databricks. Informatics teams can ingest raw unstructured medical text files into Databricks, extract meaningful insights using natural language processing techniques, and make the data available for downstream analytics. We have specific NLP solutions for extracting oncology insights from lab reports, automating the de-identification of PHI and identifying adverse drug events.

Solution Accelerator: Disease Risk Prediction

One of the most powerful tools for identifying patients at risk for a chronic condition is the analysis of real world data (RWD). This Solution Accelerator notebook provides a template for building a machine learning model that assesses the risk of a patient for a given condition within a given window of time based on a patient's encounter history and demographics information.

Demo: Real-Time COVID-19 Contact Tracing

Databricks COVID-19 surveillance solution takes a data-driven approach to adaptive response, applying predictive analytics to COVID-19 data sets to help drive more effective shelter-in-place policies.

Customer story

CRISP
Connecting Physicians With Technology to Improve Patient Care in Maryland

WATCH THE VIDEO >

From Vaccine Management to ICU Planning

During the pandemic, the Chesapeake Regional Information System for our Patients implemented a modern data architecture on Databricks to address critical reporting needs. This allowed them to analyze 400 billion data points for innovative use cases like real-time disease spread tracking, vaccine distribution and prioritizing vulnerable populations.

Conclusion

Today, data is at the core of how government agencies operate and AI is at the forefront of driving innovation into the future. The Databricks Lakehouse for Public Sector enables government agencies at the federal, state and local level to harness the full power of data and analytics to solve strategic challenges and make smarter decisions that improve the safety and quality of life of all citizens.

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

START YOUR FREE TRIAL

Contact us for a personalized demo
databricks.com/contact

databricks

Databricks is the data and AI company. More than 7,000 organizations worldwide — including Comcast, Condé Nast, H&M and over 40% of the Fortune 500 — rely on the Databricks Lakehouse Platform to unify their data, analytics and AI. Databricks is headquartered in San Francisco, with offices around the globe. Founded by the original creators of Apache Spark,™ Delta Lake and MLflow, Databricks is on a mission to help data teams solve the world's toughest problems. To learn more, follow Databricks on Twitter, LinkedIn and Facebook.