ENTERPRISE WHITE PAPER

Modernizing Anti-Money Laundering on the Databricks Data Intelligence Platform

An architecture blueprint for BSA/AML effectiveness: governed data, graph and machine learning detection, and human accountable AI for investigations.

Prepared for	Chief Compliance Officers, BSA Officers, AML Investigation Leads, Heads of Financial Crime Technology
Prepared by	Infinitive, a Databricks consulting partner, in support of the Databricks for AML solution
Delivery vehicle	AML Catalyst: modular financial crime risk management services for community, regional, and global banks
Version and date	Version 1.0, August 2026

Contents

How to read this paper

Sections 1 and 2 are written for executive sponsors and second line risk leadership. Section 3 is written for enterprise and data architects and can be read on its own as a technical blueprint. Sections 4 and 5 are written for program sponsors who need to build a business case and a defensible sequencing plan.

1. Executive Summary

AML programs face growing pressure to demonstrate that controls are risk-based, effective, and supported by reliable evidence. For many institutions, fragmented data and disconnected workflows make this difficult—even when established detection systems and investigation teams are working diligently. The Financial Crimes Enforcement Network has proposed a fundamental reform of AML/CFT program requirements that moves supervisory attention away from technical compliance and toward demonstrable, risk-based effectiveness, and it explicitly contemplates machine learning and artificial intelligence as legitimate means of improving that effectiveness.

The Wolfsberg Group has argued the same point from the industry side since 2019, framing effectiveness as compliance with law, the provision of highly useful information to authorities, and a reasonable risk-based control set rather than volume of alerts processed. [1, 6] Meanwhile, the Federal Financial Institutions Examination Council continues to place suspicious activity monitoring and reporting at the center of the examination, which means every claim of effectiveness must be evidenced with traceable data. [3]

The problem is architectural, not procedural

Many AML programs face limitations not only in detection logic, but also in the fragmented data, identity resolution, workflow context, and evidence-management capabilities surrounding it. Customer, transaction, alert, case, and investigator data are spread across core banking, know your customer platforms, sanctions screening, transaction monitoring engines, and case management, each with its own identity resolution and refresh cadence. The consequences are consistent and measurable: alert queues dominated by false positives, investigators who spend the majority of their day gathering data rather than analyzing it, and lineage that is in most cases reconstructed by hand for every examination.

The proposition

Unify customer, transaction, alert, case, and investigator data on one governed platform. Keep rules-based detection, which examiners understand and validate, and layer entity resolution, graph analytics, and machine learning risk scoring on top of it to rank the alerts those rules produce. Put a human accountable agentic layer in front of the investigator so that evidence gathering, policy retrieval, adverse media screening, and first-draft narrative writing are complete before the case is opened. Govern all of it, data and models and agent calls alike, through a single control plane so that audit readiness becomes a property of the platform rather than a quarterly project.

Target outcomes

Dimension	Typical starting point	Target after modernization
False positive burden	90 to 95 percent of alerts closed as noise	50 to 75 percent fewer alerts reaching an investigator
Case investigation time	3 to 6 hours of manual evidence gathering	10 to 20 minutes with agent-assembled evidence
SAR narrative drafting	1 to 2 hours per filing, written from scratch	Under 4 minutes to a reviewable draft
Risk scoring accuracy	Static rule thresholds, limited tuning evidence	Greater than 90 percent scoring accuracy, monitored and versioned
Investigation and reporting cost	Baseline operating expense	30 to 40 percent reduction through consolidation
Audit evidence	Assembled manually before each examination	Produced continuously from platform lineage

Table 1. Program targets used throughout this paper. Baselines are drawn from the supporting program collateral and should be re-established against institution-specific data during discovery.

The executive ask

The recommendation in this paper is not a multi-year platform commitment or a complete disposal of current systems and solutions. It is to consider the ways to augment, enhance, and identify ways to optimize investigator productivity, gain control and visibility over banking data, and enhance risk identification and mitigation.

2. The AML Landscape and Modern Challenges

2.1 The regulatory pivot toward effectiveness

In April 2026 FinCEN issued a notice of proposed rulemaking that would reform AML/CFT program requirements across covered institutions, replacing a largely technical compliance model with an effectiveness-based, risk-driven framework. Programs would be expected to rest on a documented risk assessment process, to incorporate the government-wide AML/CFT priorities, and to direct attention and resources toward higher-risk customers and activities rather than lower-risk ones. [1, 2] For technology leadership the implication is direct. A program that cannot show which risks drove which controls, which data fed which model, and which evidence supported which disposition cannot demonstrate effectiveness, however many alerts it clears.

Three further reference points frame the target state. The FFIEC BSA/AML Examination Manual continues to treat suspicious activity monitoring and reporting as a core control, which sets the evidentiary bar. [3] The FATF Recommendations and the FATF assessment of new technologies establish the risk-based approach internationally and recognize that analytics can strengthen it rather than dilute it. [4, 5] And the 2018 interagency joint statement on innovative efforts, together with supervisory guidance on model risk management and its 2021 BSA/AML specific interagency statement, establishes that supervisors expect innovation to be accompanied by validation, documentation, and governance rather than replaced by it. [8, 9, 10]

2.2 Typologies have outrun single-system detection

Laundering typologies have adapted faster than the platforms built to catch them. The common thread across the patterns below is that none of them are visible inside a single account, a single product, or a single day of activity. They are visible in relationships.

Typology	How it appears in the data	Why single-system detection misses it	What closes the gap
Structuring and smurfing	Many deposits or withdrawals held below reporting thresholds across accounts, branches, and days.	Threshold rules evaluate one account at a time and have no view of coordinated behavior across parties.	Motif finding across a transaction graph plus behavioral baselining at the customer and network level.
Layering and round tripping	Funds move through intermediaries and return to origin, often within	Each hop is individually unremarkable, and counterparty data usually	Connected components and path analysis on a unified graph, with risk

	days, with no economic purpose.	sits in a different system.	propagated across the network.
Synthetic identity	Fabricated customers assembled from real and invented attributes that share addresses, devices, or contact details.	Deterministic matching treats near duplicates as distinct people, so the network is never assembled.	Probabilistic entity resolution at the silver layer, before detection logic is applied.
Funnel accounts and mule networks	Deposits in many geographies followed by concentrated withdrawal, high account turnover, short lifetime.	Geographic and channel data are fragmented, and case history is not reusable across investigations.	Conformed customer and counterparty model plus reuse of prior case evidence in retrieval.
Trade-based laundering	Invoice values inconsistent with shipment, goods, or corridor norms, evidenced largely in documents.	Documents are unstructured and sit outside the monitoring engine entirely.	Document ingestion into governed volumes, indexed for semantic retrieval alongside structured activity.
Rapid rails abuse	Instant payments and faster settlement compress the window between initiation and irreversibility.	Batch scoring cycles evaluate activity long after the funds have moved.	Streaming ingestion with real-time model serving on the same governed data used for batch.

Table 2. Priority typologies and the detection capabilities they require.

2.3 Where legacy vendor platforms reach their limit

This is not an argument for removing established detection vendors. Rules engines from established platforms carry validation history and examiner familiarity that has real value, and the architecture in Section 3 is designed to read from them rather than replace them. The limits show up in the layers around detection.

- **Closed data models.** Alerts, features, and dispositions are held inside a proprietary schema, so enterprise analytics and model risk teams work from extracts rather than from the system of record.
- **Batch-bound processing.** Nightly or intraday cycles were designed for a settlement world, not for instant payments.
- **Weak identity resolution.** Deterministic matching leaves the same beneficial owner represented several times, which fragments the network before any graph analysis can be attempted.
- **Tuning as a project.** Threshold changes require vendor engagement and long validation cycles, so thresholds drift from the risk profile they were meant to reflect.
- **Unstructured data left outside.** Policies, prior narratives, correspondence, and adverse media are not part of the detection or investigation surface.

- **Lineage reconstructed by hand.** The evidence examiners request exists, but assembling it is a manual exercise repeated at every examination.
- **Compounding license cost.** Separate warehouses, graph tools, machine learning platforms, and business intelligence licenses accumulate around the detection engine.

The reframing that matters

Most banks do not need one more AML point solution. They need one governed intelligence layer that every control workflow, including the detection vendors already in place, can read from and write into.

3. Architectural Blueprint on Databricks

The blueprint below has four layers: a governed data foundation, an AML-optimized data model, detection and analytics, and a human accountable agentic layer, with unified governance running through all of them. Each layer is independently deployable. A bank can start with any one of the three delivery modules described in Section 5 and still converge on the same target architecture.

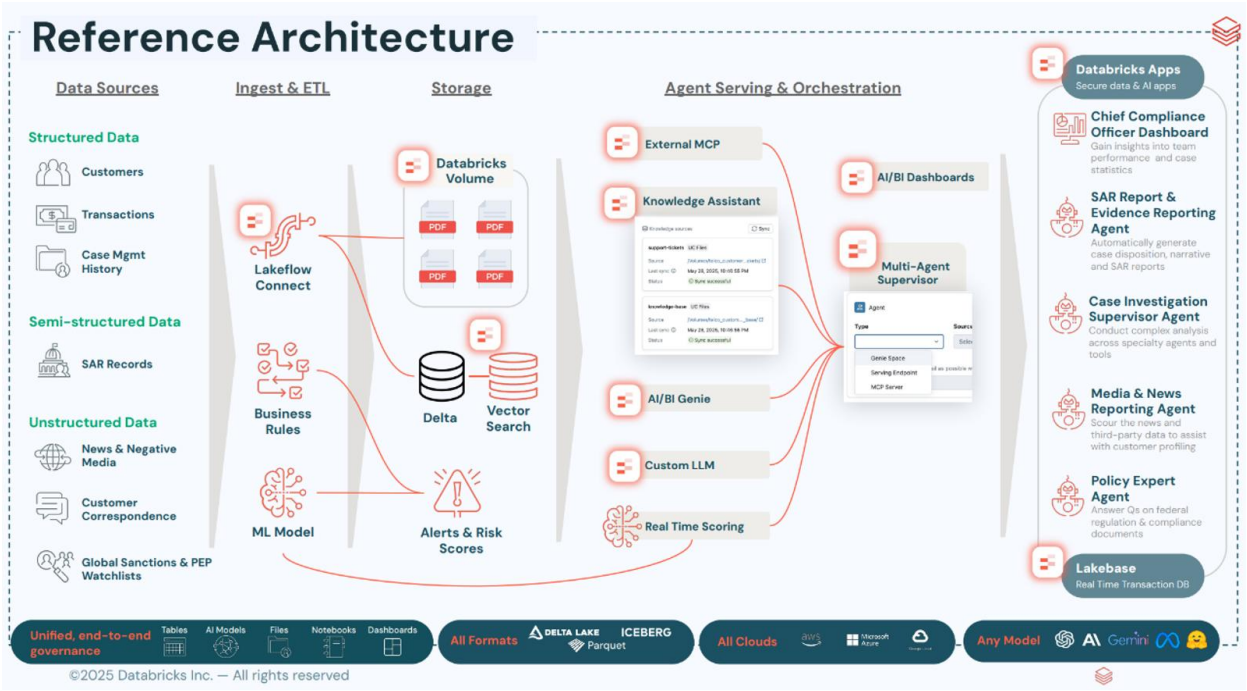

Figure 1. End-to-end AML reference architecture on the Databricks Data Intelligence Platform.

3.1 Ingestion and storage

Federation before migration (where possible)

The fastest route to a governed foundation is usually not always data migration. Unity Catalog Federation registers existing warehouses, including Teradata, Oracle, and SQL Server, so that access policy, semantics, and lineage apply to data that has not moved. That distinction matters to sequencing: governance and unified query arrive in weeks, while physical consolidation proceeds only where it earns its cost. There are dependencies on sources that are supported, proper configuration, permission settings, and the institution's architecture.

Lakeflow Connect for batch and streaming

Where data does need to land, Lakeflow Connect can support batch and streaming ingestion on one platform: core banking extracts, wire and ACH feeds, sanctions and

watchlist files, screening output, and alert and disposition records from existing monitoring engines. Properly developed streaming pipelines carry the payment flows where detection latency has real consequences, and the same declarative pipeline definitions serve both modes, which removes the classic divergence between a real-time path and a batch path that score differently.

Delta Lake as the storage substrate

Delta Lake provides the properties an AML system of record requires: ACID transactions so that concurrent ingestion and investigation reads never expose partial state, schema enforcement and evolution so that a new source field cannot silently corrupt a detection feature, and time travel so that an analyst or examiner can reconstruct exactly what the data looked like on the day a disposition was made. That last property is the difference between asserting a decision was reasonable and demonstrating it.

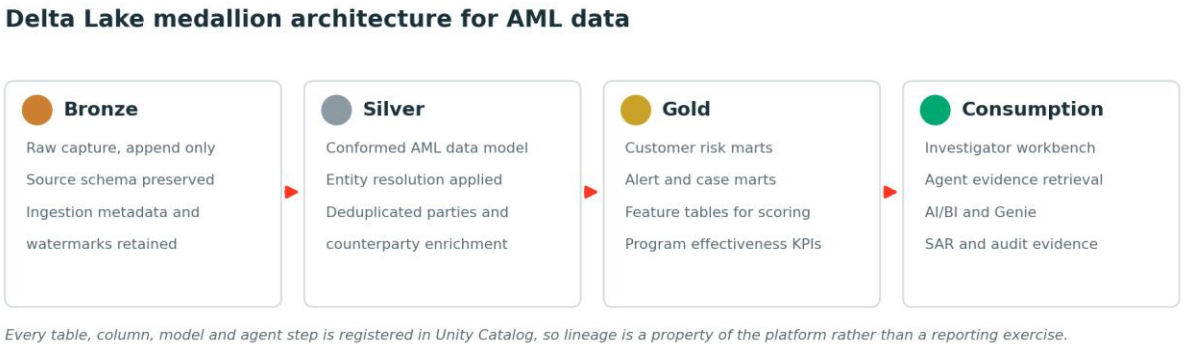

Figure 2. Medallion layering applied to AML data, with entity resolution performed at the silver layer.

- **Bronze.** Raw, append-only capture with source schema and ingestion metadata preserved, so the original record is always recoverable.
- **Silver.** The conformed AML data model: parties, accounts, transactions, counterparties, alerts, cases, and watchlist hits, with entity resolution and deduplication already applied.
- **Gold.** Purpose-built marts for customer risk, alert triage, case management, and program effectiveness reporting, plus feature tables consumed by detection models.
- **Volumes and Databricks Vector Search Indexes.** Policy documents, regulatory guidance, prior SAR narratives, and correspondence stored in governed Unity Catalog volumes and indexed for semantic retrieval.
- **Lakebase.** Serverless transactional storage for real-time case state and application interactions, sitting beside the analytical tables rather than in a separate operational silo.

3.2 Detection and analytics

Entity resolution as a prerequisite, not a feature

No graph or model is better than the identities underneath it. Probabilistic entity resolution reconciles parties across source systems using name, address, date of birth, device, and contact attributes, producing stable identifiers and a confidence score that downstream logic can reason about. The published Databricks AML solution accelerator demonstrates this pattern using open source linkage libraries alongside graph analysis, and it is the correct place to spend effort first because every later capability inherits its quality.

Graph analytics with GraphFrames

With resolved entities, the transaction graph becomes tractable. GraphFrames on Spark supports the three operations that matter for laundering typologies at scale.

- **Connected components.** Assemble the network of accounts, devices, addresses, and beneficial owners that belong together, which is how synthetic identity clusters and shell structures surface.
- **Motif finding.** Express structuring, round tripping, and fan-in fan-out patterns as graph patterns and match them across billions of edges, rather than approximating them with account-level thresholds.
- **Risk propagation.** Use iterative message passing to spread risk from a confirmed bad node across the network, so exposure to a known typology raises the score of its neighbors.

These are the exact techniques published in the Databricks AML solution accelerator, which combines graph analytics, natural language processing on counterparty data, and machine learning scoring on the lakehouse, with detected patterns persisted back into Delta tables for SQL analysis and investigator consumption. [12, 13, 14]

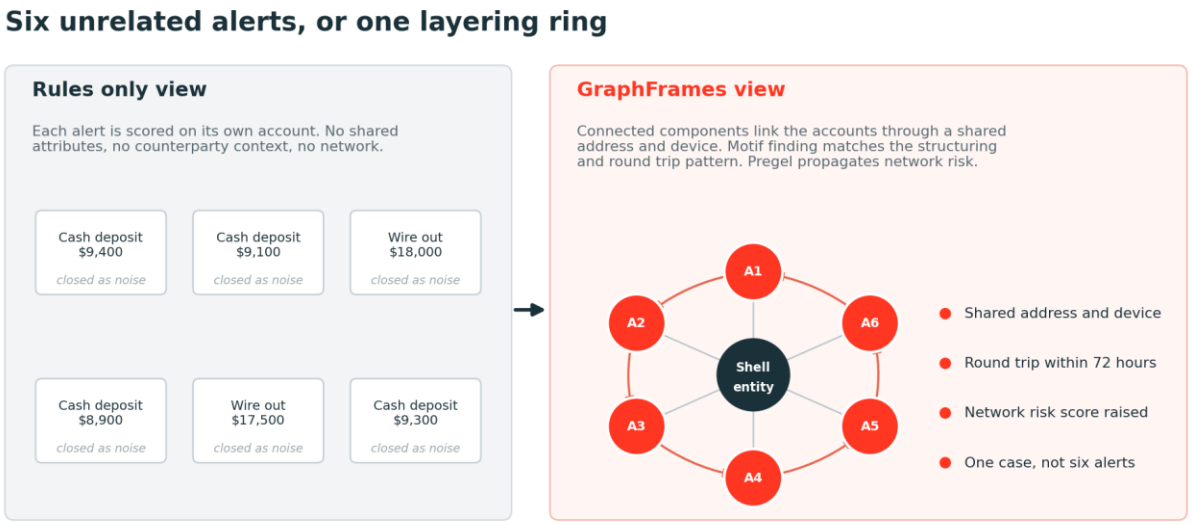

Figure 3. The same six alerts seen by a rules-only engine and by graph analytics on resolved entities.

Machine learning risk scoring on top of rules, not instead of them

Rules are retained. They are explainable, examiner-validated, and cheap to reason about. The machine learning layer does not decide whether an alert exists. It ranks the alerts the rules produce, on a 0 to 100 scale, using features that the rules cannot see: network position, peer group deviation, behavioral drift against the customer's own baseline, counterparty risk, and document and adverse media signal. Investigators work the top of the queue; low-scoring alerts are dispositioned with an auto-generated rationale that a human still approves. This is the mechanism behind the 50 to 75 percent reduction in alerts reaching an investigator, and it is deliberately structured so that no alert is suppressed by a model without a recorded, reviewable basis.

- **Unsupervised anomaly detection.** Isolation forest and autoencoder approaches surface behavior with no labeled precedent, which is where new typologies first appear.
- **Supervised risk ranking.** Gradient boosted models trained on confirmed SAR outcomes and quality-controlled dispositions, retrained on a governed cadence.
- **Sequence and peer models.** Deviation from a customer's own history and from a dynamically constructed peer group, which is more durable than static segment thresholds.
- **Registered and monitored.** Every model version, training dataset, evaluation result, and deployment is recorded in MLflow and surfaced to model risk management.

3.3 The agentic investigation layer

The largest single cost in an AML operation is not detection. It is the three to six hours an investigator spends assembling context that already exists somewhere in the institution. A multi-agent system built on Agent Bricks removes that work without removing the investigator from the decision.

Agent	Responsibility	Grounding and control
Evidence Gatherer	Assembles transaction history, KYC profile, prior alerts and cases, beneficial ownership graph, and sanctions context before the case is opened.	Reads only governed tables the investigator is entitled to see, under Unity Catalog access policy.
Policy Expert	Answers questions against the institution's red flag library, standard operating procedures, and federal guidance.	Retrieval augmented generation over policy documents indexed in Vector Search Indexes, with citations back to source.
Risk Reasoner	Produces a structured rationale for escalation or closure that the analyst can accept, edit, or reject.	Structured output with named contributing factors, never an unexplained score.
SAR Drafter	Generates a first-pass narrative grounded in the case evidence and formatted to filing expectations.	Draft only. A named human reviews, edits, and signs off before anything is filed.
Adverse Media Researcher	Screens negative news and external sources in real time to enrich the case picture before triage.	Governed external search through a controlled connector, with retrieved sources retained as evidence.
Supervisor	Routes work across the specialist agents and assembles the case package.	Every tool call, retrieved document, and intermediate step traced in MLflow.

Table 3. Specialist agents in the investigation layer. Human sign-off is a control, not a courtesy.

Model choice is a governed configuration rather than an architectural commitment. Agent calls are routed through the Unity Gateway. Unity Gateway can centralize model access, policy controls, observability, and cost management; institutions should apply their established validation and change-management processes to material model changes. Compliance leadership consumes the same governed data conversationally through AI/BI Genie, which answers operational questions such as which analyst is carrying alerts past their service level, without a request to the business intelligence team.

Human accountable AI, not AI everywhere

Agents gather, retrieve, structure, and draft. Humans decide. Every escalation, closure, and filing carries a named accountable reviewer, and every input that reviewer saw is retained as evidence. That posture is what makes the productivity gain defensible under

examination.

3.4 Governance, lineage, and explainability

Unity Catalog is the single control plane for data and artificial intelligence assets: one permission model across tables, files, features, models, and agent endpoints; row-level and column-level controls so that investigator entitlement is enforced at the data rather than in application code; automatic end-to-end lineage from source column through feature and model to the dashboard or agent output; and audit logging of access and change. Supervisory guidance on model risk management, including the 2021 interagency statement addressed specifically to systems supporting BSA/AML compliance, expects exactly this combination of documentation, validation evidence, and change control. [9, 10]

- **Lineage as evidence.** The question of which data supported a decision is answered by a query against catalog metadata rather than by a reconstruction project.
- **Lakehouse Monitoring as a control.** Data quality expectations, feature drift, score distribution shift, and agent evaluation results are monitored continuously, with alerting when a metric leaves tolerance. Continuous monitoring is subject to the deployment and configuration of the relevant monitors and alerts.
- **Explainability by construction.** Rules carry their own logic. Machine learning scores carry ranked contributing factors and reason codes suitable for narrative use. Agents carry a full trace of retrieved sources and tool calls.
- **Model risk alignment.** Development documentation, validation results, ongoing performance monitoring, and version history are produced by the platform rather than assembled for each validation cycle.
- **Separation of duties.** Development, validation, and production promotion are distinct roles with distinct entitlements, recorded in the catalog.

Explainability deserves a specific note. A score that cannot be explained cannot be used in a filing, and a model that cannot be validated will not survive examination. The design rule applied throughout this blueprint is that any output that touches a regulatory decision must carry its basis with it: the rule that fired, the features that drove the score, the documents the agent retrieved, and the human who approved the result.

4. Business Impact and Case Benchmarks

The benchmarks below are program targets from the supporting collateral, expressed so that each one is measurable during a pilot rather than only at the end of a program. Every institution should re-baseline them during discovery, because the value of a modernization is the delta from the institution's own starting point.

Metric	Typical baseline	Target	How it is evidenced
False positive rate	90 to 95 percent	50 to 75 percent fewer alerts reaching an investigator	Alert volume and disposition tracked before and after, with suppressed alerts sampled and reviewed
Alert to SAR ratio	Low and unstable	Materially improved with risk-ranked queues	Ratio reported by typology and by scoring band
Average investigation time	3 to 6 hours per case	10 to 20 minutes per case	Case timestamps captured in the workbench, not self-reported
SAR drafting time	1 to 2 hours per filing	Under 49 minutes to reviewable draft	Draft generation timestamp to reviewer sign-off
Risk scoring accuracy	Not consistently measured	Greater than 90 percent	Held-out evaluation against confirmed outcomes, recorded in MLflow
Agent response time	Not applicable	Under 5 seconds for investigator queries	Endpoint latency monitoring
Investigation and reporting cost	Baseline operating expense	30 to 40 percent reduction	Cost per case and per filing, plus retired tooling spend
Examination preparation effort	Weeks of manual assembly	Continuous, query-based evidence	Time to produce a full decision lineage on request

Table 4. Outcome targets and the evidence that substantiates each one.

4.1 Where the operational savings come from

- **Triage cost.** The dominant saving is not faster investigation of every alert. It is the alerts that never reach an investigator because a risk score, supported by a recorded rationale, placed them at the bottom of the queue.
- **Evidence assembly.** Data gathering compresses from an hour or more per case to effectively zero, because the evidence package is built before the case is opened.

- **Narrative production.** Reusable language and grounded drafting remove the from-scratch cost of each filing while leaving editorial control with the investigator.
- **Platform consolidation.** Overlapping warehouses, graph tools, machine learning platforms, and business intelligence licenses collapse into one platform, which is the source of the 30 to 40 percent operating cost reduction.
- **Examination readiness.** Preparation stops consuming senior compliance capacity, which is a real cost even though it rarely appears in a program business case.

4.2 Illustrative value model

The following model is illustrative and uses round numbers to show the shape of the arithmetic, not a forecast. Substitute institutional figures during discovery.

Assumption or result	Value	Basis
Alerts per year	120,000	Institution baseline
Alerts reaching an investigator today	120,000	No risk-based triage in place
Average handling time per alert	3.5 hours	Midpoint of the 3 to 6 hour range
Analyst hours per year on alerts	420,000	Volume multiplied by handling time
Alert volume reduction at the conservative end	50 percent	Risk-ranked queue with recorded rationale
Handling time reduction on remaining alerts	70 percent	Agent-assembled evidence and drafting
Analyst hours after modernization	63,000	60,000 alerts at approximately 1.05 hours
Hours released	357,000	Redeployed to higher-risk work and backlog reduction

Table 5. Illustrative arithmetic only. The purpose is to identify which variable matters most: the alert volume reaching an investigator.

Read the model correctly

Released capacity is not automatically a headcount reduction, and framing it that way tends to damage adoption. In most programs the first and most defensible use of released capacity is clearing backlog, deepening high-risk investigations, and expanding typology coverage that the institution has been deferring.

5. Implementation Roadmap: Phased Migration from Legacy Systems

The sequencing principle is straightforward: govern first, prove value in one workstream, and decommission only what the new environment has demonstrably replaced. Nothing in this roadmap requires a full data migration before the first outcome is delivered, because federation gives governed access to legacy warehouses in place.

5.1 Choose the entry module

Module	Start here when	Core build	Primary KPIs
KYC and Due Diligence	Customer context is the bottleneck and beneficial ownership is unreliable.	Unify KYC data, map to the AML data model, graph-based entity resolution, machine learning risk assessment.	Profile accuracy, percentage of ultimate beneficial owners identified, documentation exception rate.
AML Alerting and Monitoring	False positive volume is the dominant pain and the backlog is growing.	Tunable rule-based detection on governed data, machine learning risk scoring, alerting dashboards.	False positive rate, alert to SAR ratio, alert volume trend, suppression rate.
Investigate and Report	Detection is acceptable but case handling and filing timeliness are not.	Investigation assistant agents, investigator workbench, AI-generated SAR drafts, governed external search.	Average time to close, open alert count, filing timeliness, investigation quality score.

Table 6. The three delivery modules. Each is independently deployable and each converges on the same target architecture.

5.2 Phased plan

Phase	Indicative duration	Key activities	Exit criteria
0. Architecture and transformation review	1 to 2 weeks	Current-state mapping of data, workflows, controls, and handoffs. Gap analysis across process, data, technology, and governance. Module selection.	Agreed problem statement, success criteria, executive sponsor, and a prioritized module.
1. Discovery and data foundation	2 to 4 weeks	Source inventory across monitoring, KYC, and core banking. Schema mapping to the AML data model. Unity	Functional requirements documented, data model mapping complete, governance model

		Catalog and federation setup. Requirements sign-off.	defined.
2. Platform deployment	3 to 6 weeks	Build or federate pipelines. Configure detection logic and risk scoring. Deploy the application layer, real-time store, and agents.	Pipelines operational, detection logic configured, application and agents deployed.
3. Pilot and refinement	2 to 4 weeks	Pilot cohort rollout, user acceptance testing, parallel run against the incumbent engine, tuning of scoring and agent prompts, feedback capture.	User acceptance testing complete, pilot sign-off, performance and latency within tolerance.
4. Production rollout	2 to 4 weeks	Production alert monitoring and alerting, go-live support, change management and training, control testing.	Production go-live, training complete, control evidence captured in the workflow.
5. Scale and decommission	Ongoing	Add the next module, retire overlapping tooling once parallel run evidence supports it, expand typology coverage.	Documented decommissioning decisions with validation evidence for each retired control.

Table 7. Phased path to production. Community, regional, and global institutions typically run 6 to 8, 10 to 12, and 12 to 16 weeks per module respectively.

5.3 Migration and parallel run discipline

- **Never cut over on a promise.** Run the new scoring alongside the incumbent engine on the same population and compare dispositions before any threshold is retired.
- **Sample what you suppress.** Alerts that the risk score deprioritizes are sampled and reviewed on a defined cadence, and the sampling result is part of the tuning evidence.
- **Validate before you rely.** Every model that influences a regulatory decision enters the institution's model risk management process with development documentation, validation results, and alert monitoring plans.
- **Retire deliberately.** Each decommissioned control gets a written record of what replaced it and what evidence supports the replacement.
- **Keep the rules.** Rules-based detection remains the explainable backbone. The modernization changes what surrounds it, not whether it exists.

5.4 The delays that do not appear on the project plan

Programs of this kind rarely stall on engineering. They stall on the operating model, and the failure modes are predictable enough to plan against.

- **Adoption that never lands.** If investigators do not trust agent-generated evidence or the new triage logic, they revert to old habits and the institution pays for a new platform while running the old process on top of it. The countermeasures are role-based training, governance forums where reviewers pressure-test outputs before depending on them, phased cohort rollout, and measuring usage and quality rather than deployment.
- **Controls that live on paper.** Controls that depended on manual steps break silently when the workflow changes. For every critical step in the new operating model, define ownership, evidence requirements, escalation paths, and testing standards, and capture lineage and evidence automatically rather than reconstructing it later.
- **Building on broken assumptions.** Inconsistent lineage, weak entity resolution, manual evidence collection, and unclear ownership of model tuning each widen once code is built on top of them. A structured gap analysis across process, data, technology, and governance belongs before the first pipeline, not after the pilot.

Wrap every module the same way

Impact assessment, regulatory review and control alignment, and gap analysis apply to each module, alongside organizational change management: strategy and roadmap, analyst and investigator readiness, communication planning, and resistance management. The deliverable is not only working software. It is examiner-ready documentation and an operating model the institution can sustain.

6. Conclusion and Recommended Executive Action

The regulatory direction of travel, the economics of alert handling, and the technical maturity of governed lakehouse architecture now point the same way. A United States bank cannot materially improve AML effectiveness while customer, transaction, alert, case, and investigator data remain in separate systems with separate identity resolution. Unifying that data under one governance plane, keeping explainable rules while adding entity resolution, graph analytics, and machine learning risk ranking, and putting a human accountable agentic layer in front of the investigator is a change to the operating model that the control environment can absorb, because every step of it produces its own evidence.

The recommended executive action is approve a time-boxed discovery and pilot designed around one priority workflow. The duration and scope should reflect the institution's data readiness, control requirements, integration complexity, and validation process. Scale only after the institution has reviewed measurable impact, control evidence, and operating-model readiness.

1. Sponsor alignment. Confirm the executive sponsor, the problem statement, and the success criteria.
2. Diagnostic sprint. Map current data, workflows, controls, and handoffs across AML operations.
3. Blueprint and business case. Prioritize the lighthouse module and define the target-state architecture.
4. Governed foundation. Stand up the initial AML data products, access model, and Lakehouse Monitoring.
5. First proof point. Deliver one module and a measured detection or investigation improvement.
6. Decision gate. Approve scale-up only once impact, control posture, and operating model are proven.

What is being approved

Discovery, one lighthouse module sized to the institution's profile to be proven working alongside existing tech-stack to create demonstrable impact and business justification for further implementation identifying KPI targets for cost savings, productivity gains, risk reduction with a clear decision gate. Not a multiyear platform commitment.

7. Notes and References

Superscript markers in the body text refer to the numbered sources below. Sources are limited to primary regulatory bodies, recognized industry standard setters, and official platform documentation.

1. Financial Crimes Enforcement Network. Anti-Money Laundering and Countering the Financing of Terrorism Programs. Notice of Proposed Rulemaking, April 7, 2026, published in the Federal Register April 10, 2026.
2. Financial Crimes Enforcement Network. Fact Sheet: Proposed Rule to Fundamentally Reform Financial Institution AML/CFT Programs, April 2026.
3. Federal Financial Institutions Examination Council. Bank Secrecy Act / Anti-Money Laundering Examination Manual, sections on suspicious activity reporting and transaction monitoring.
4. Financial Action Task Force. International Standards on Combating Money Laundering and the Financing of Terrorism and Proliferation, the FATF Recommendations, as amended.
5. Financial Action Task Force. Opportunities and Challenges of New Technologies for AML/CFT, July 2021.
6. The Wolfsberg Group. Statement on Effectiveness, 2019; Developing an Effective AML/CTF Programme, 2020; Demonstrating Effectiveness, 2021.
7. The Wolfsberg Group. Statement on Effective Monitoring for Suspicious Activity, Part I: Moving Beyond Automated Transaction Monitoring, 2024.
8. Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, National Credit Union Administration, Office of the Comptroller of the Currency, and Financial Crimes Enforcement Network. Joint Statement on Innovative Efforts to Combat Money Laundering and Terrorist Financing, December 3, 2018.
9. Office of the Comptroller of the Currency, Bulletin 2011-12, and Board of Governors of the Federal Reserve System, SR 11-7. Supervisory Guidance on Model Risk Management.
10. Board of Governors of the Federal Reserve System, FDIC, NCUA, OCC, and FinCEN. Interagency Statement on Model Risk Management for Bank Systems Supporting BSA/AML Compliance, April 9, 2021.
11. Financial Conduct Authority. Financial Crime Guide: A Firm's Guide to Countering Financial Crime Risks.
12. Databricks. Solution Accelerator: Compliance, AML and KYC. Official Databricks solution accelerator library.
13. Databricks. AML Solutions at Scale Using the Databricks Lakehouse Platform, engineering blog, July 16, 2021.
14. Databricks Industry Solutions. Anti-Money Laundering accelerator repository, covering GraphFrames motif finding, connected components, entity linkage, and Delta Lake persistence of detected patterns.
15. Databricks product documentation for Unity Catalog, Unity Catalog Federation, Lakeflow Connect, Delta Lake, Vector Search Indexes, Model Serving, MLflow, Agent Bricks, Unity Gateway, AI/BI and Genie, and Lakebase.

Appendix A. Control to Evidence Mapping

The mapping below is the artifact examiners and internal audit ask for most often. It is the practical answer to the question of how a modernized program evidences the controls it claims.

Control objective	Platform capability	Evidence artifact
Complete and accurate AML data	Delta Lake with schema enforcement, pipeline expectations, and quality monitoring	Pipeline run history, expectation pass rates, quality dashboards
Customer identity is resolved consistently	Probabilistic entity resolution at the silver layer with confidence scoring	Match rules, confidence distributions, manual override log
Detection is risk-based and tuned	Rules plus machine learning risk scoring with versioned thresholds	Threshold change history, tuning analysis, above and below the line testing
Models are validated and monitored	MLflow registry, evaluation suites, drift and performance monitoring	Model documentation, validation results, monitoring alerts, version history
Access is appropriate and least privilege	Unity Catalog row and column level policy across data, models, and endpoints	Entitlement reports, access review evidence, audit logs
Decisions are traceable end to end	Automatic lineage from source column to model to output	Lineage graph for any alert, case, or filing on request
AI use is governed	Unity Gateway routing, permissions, and Lakehouse Monitoring for observability, with MLflow tracing of agent steps	Agent interactions and relevant traces can be captured and monitored using configured observability and governance controls
Filings are human accountable	Draft-only agent output with named reviewer sign-off	Reviewer identity, edit history, and approval timestamp per filing

Table 8. Control objectives mapped to platform capability and the evidence produced.

Appendix B. Metric Definitions

Metric	Definition
False positive rate	Alerts closed without escalation as a proportion of all alerts generated in the period.
Alert to SAR ratio	Filed suspicious activity reports as a proportion of alerts generated, reported by typology.

Alert suppression rate	Proportion of alerts deprioritized by risk score below the investigator review threshold, with sampling coverage reported alongside.
Average time to close	Elapsed working time from case assignment to final disposition, captured from workbench timestamps.
Investigation quality score	Quality assurance review outcome against a defined rubric covering evidence sufficiency, narrative quality, and policy citation.
Risk scoring accuracy	Model performance against held-out confirmed outcomes on the metric agreed with model risk management.
Agent response time	Elapsed time from investigator query to complete agent response at the serving endpoint.
Cost per case	Fully loaded investigation and reporting operating cost divided by cases closed in the period.

Table 9. Metric definitions to agree with second line risk before the pilot begins.