EXECUTIVE E-BOOK

Modernizing Anti-Money Laundering on databricks

Governed data, graph and machine learning detection, and human accountable AI for BSA/AML effectiveness. Delivered as AML Catalyst by Infinitive on Databricks Data + AI Platform.

90 to 95%
alerts closed without escalation

3 to 6 hrs
manual work per case today

50 to 75%
fewer alerts to investigators

Under 4 min
to a reviewable SAR draft

Illustrative pilot targets; results vary by institution, data quality, typology, operating model, and control design.

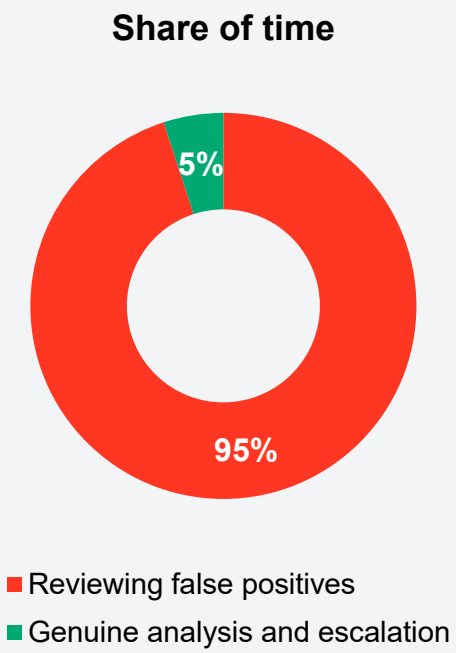
databricks × INFINITIVE

WHERE THE MONEY GOES

The cost of compliance is a data problem

Legacy bottlenecks show up as alert volume, investigator time, and evidence that has to be rebuilt.

An investigator's day



What that costs the program

- **Fragmented data.** AML context sits in core banking, KYC, screening, monitoring, and case systems, each with its own identity resolution.
- **Alert overload.** Rules fire without cross-domain context, a significant share of investigator time may be spent resolving lower-value alerts and assembling context.
- **Scavenger hunt casework.** Three to six hours per case is spent gathering data across ten or more systems.
- **Filings written from scratch.** Most narrative language is reusable, but almost none of it is reused.
- **Evidence rebuilt for every exam.** Lineage, access, and change history are assembled by hand instead of queried.

FROM POINT SOLUTIONS TO ONE INTELLIGENCE LAYER

The governed data + AI foundation for BSA/AML.

Most banks do not need one more AML tool. They need one governed layer every control workflow can trust.

TODAY

Fragmented by design

- Ten or more systems, ten or more data models
- Batch cycles built for a settlement world
- Deterministic matching splits one owner into many
- Documents and adverse media left outside detection
- Tuning is a vendor project, not a control
- Lineage reconstructed by hand at every exam

TARGET STATE

One governed intelligence layer

- Unified governed view, federated or ingested
- Streaming and batch on the same definitions
- Entity resolution before detection logic runs
- Documents indexed beside structured activity
- Rules retained, machine learning ranks the queue
- Lineage and audit evidence produced continuously

LAYER 1

Real-time ingestion and the Delta Lake core

Lakeflow handles batch and streaming on one platform. Delta Lake gives the program a governed, reliable foundation for AML data products and analytics.

Bronze

Raw capture, append only

Source schema preserved

Ingestion metadata retained

Silver

Conformed AML data model

Entity resolution applied

Counterparty enrichment

Gold

Customer and case marts

Feature tables for scoring

Effectiveness KPIs

Consumption

Investigator workbench

Agent evidence retrieval

AI/BI, Genie, and audit

ACID transactions

Investigation reads never see partial ingestion state

Schema enforcement

A new source field cannot silently break a detection feature

Time travel

Reconstruct the exact data behind any past disposition

Open formats

Delta Lake and Iceberg, on any cloud, no lock-in

LAYER 2

Potential Result...

Investigators can assess related activity as a network, rather than isolated events.

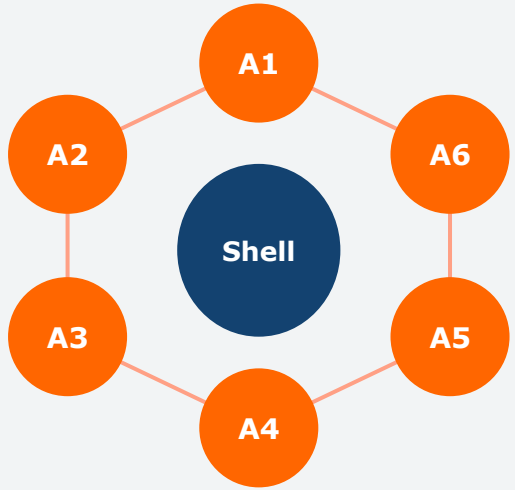
Rules only view

Each alert is scored on its own account. No shared attributes, no counterparty context.

Cash deposit $9,400 *closed as noise*	Cash deposit $9,100 *closed as noise*	Wire out $18,000 *closed as noise*
Cash deposit $8,900 *closed as noise*	Wire out $17,500 *closed as noise*	Cash deposit $9,300 *closed as noise*

GraphFrames view

Connected components, motif finding, and risk propagation on resolved entities.



Connected components. Assemble the network.

Motif finding. Match structuring and round trips.

Risk propagation. Spread score to neighbors.

Result. One case, not six alerts.

*Illustrative scenario, dependent on institution requirements

LAYER 3

AI-assisted investigations, with explainability

Agents gather, retrieve, structure, and draft. Humans decide. Configured traces and source references can support review and auditability

1 Evidence Gatherer

Transactions, KYC, prior cases, ownership, sanctions

2 Policy Expert

Red flag library, procedures, and federal guidance

3 Risk Reasoner

Structured rationale to accept, edit, or reject

4 SAR Drafter

First-pass narrative grounded in case evidence

5 Adverse Media Researcher

Governed external search before triage, subject to institution-approved sources, workflow controls, and retention requirements.

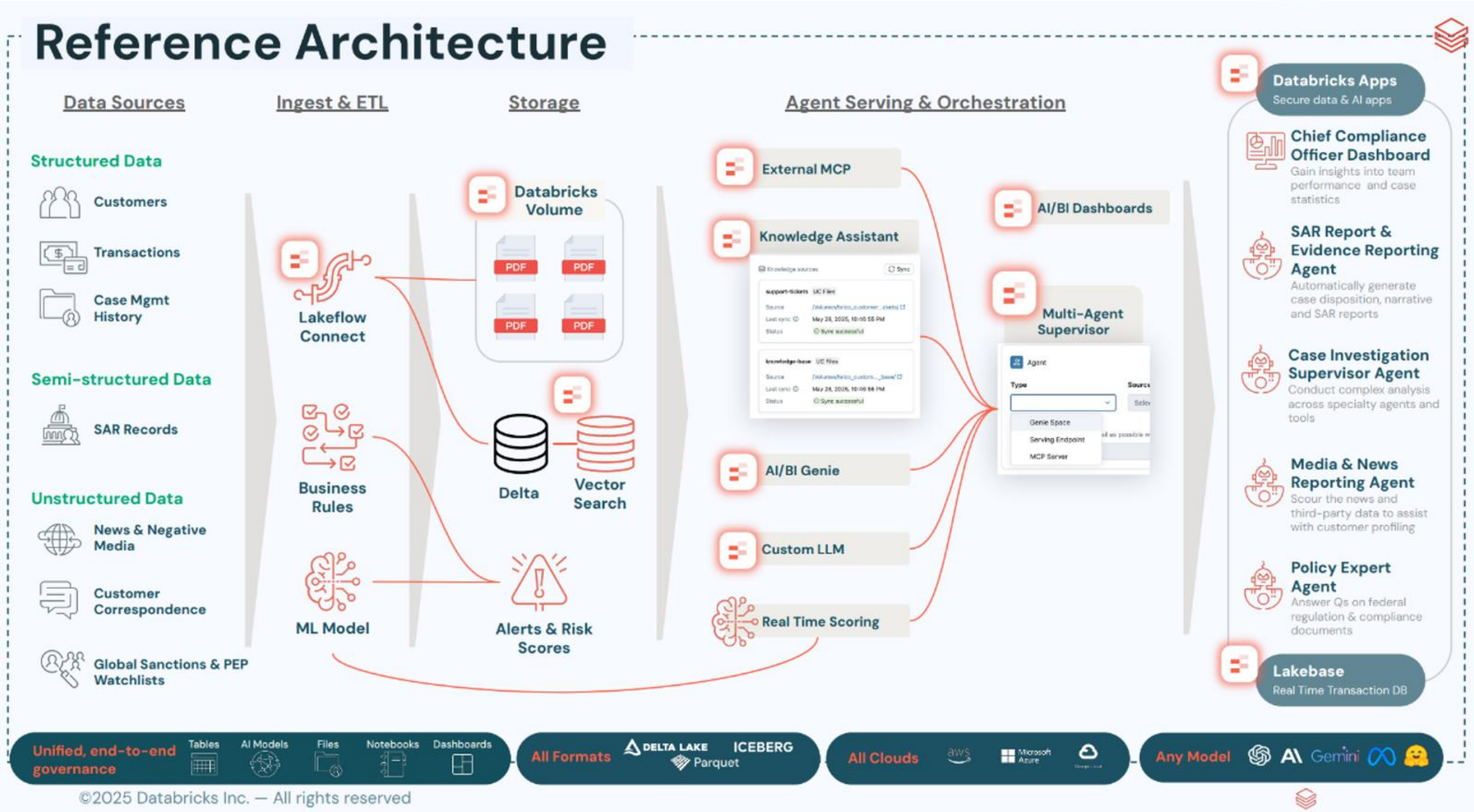
Human accountable AI, not AI everywhere

Draft only. A named reviewer signs off before anything is filed.

Explainability by construction

Rules carry their logic, scores carry ranked contributing factors, and MLflow traces every tool call.

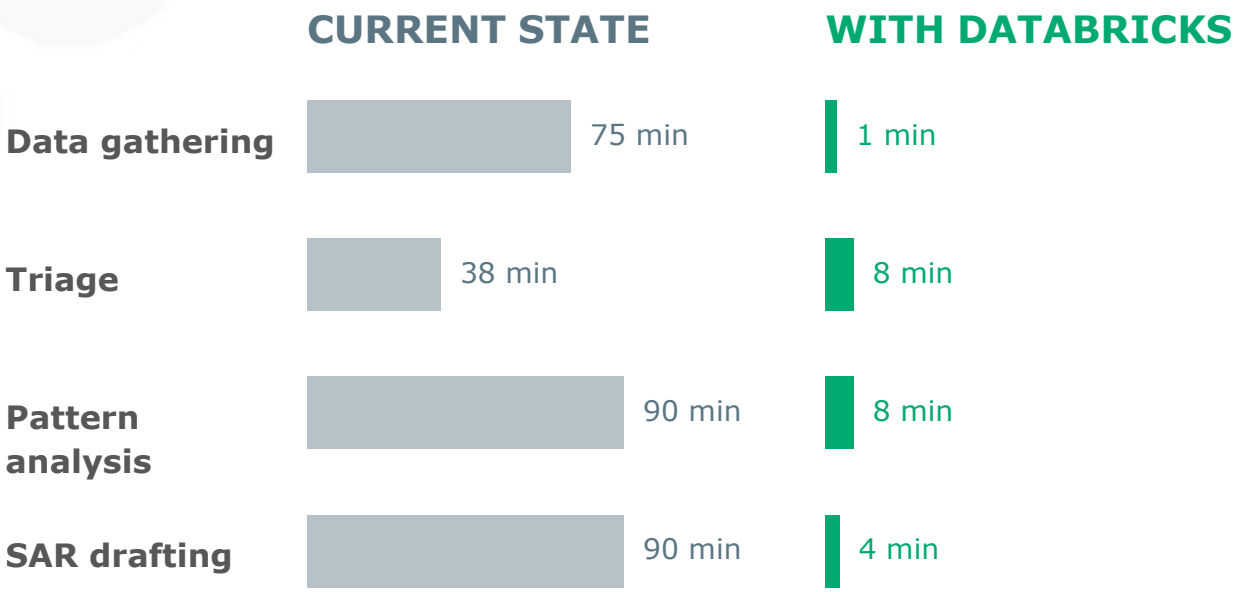
HOW THE LAYERS FIT TOGETHER



WHAT CHANGES, AND HOW IT IS MEASURED

Enterprise business outcomes

Every target below is measurable during a pilot, not only at the end of a program.



Minutes per case, by investigation stage.

50 to 75%
fewer false positives

10 to 20 min
per investigation, from 3 to 6 hours

Over 90%
risk scoring accuracy, monitored

30 to 40%
lower AML operating expense

Plus audit readiness produced continuously, not assembled per exam. Illustrative targets for validation in a pilot.

databricks × INFINITIVE

THE EXECUTIVE ASK

Approve the first 90 days

Not a multiyear program. A disciplined proof with clear decision gates.

1 Sponsor alignment

Executive sponsor, problem statement, success criteria

2 Diagnostic sprint

Map data, workflows, controls, and handoffs

3 Blueprint and case

Prioritize the lighthouse module and target state

4 Governed foundation

Initial AML data products, access model, monitoring

5 First proof point

One module and a measured, evidenced improvement

6 Decision gate

Scale once impact and control posture are proven

One governed AML intelligence platform, many control workflows, and a safer path to AI-enabled compliance productivity.

databricks × INFINITIVE