Signal Factory LLD: Out-of-Band Data Observability & Prevention

Pattern 1: External Enforcement for Immutable Producers

Objective

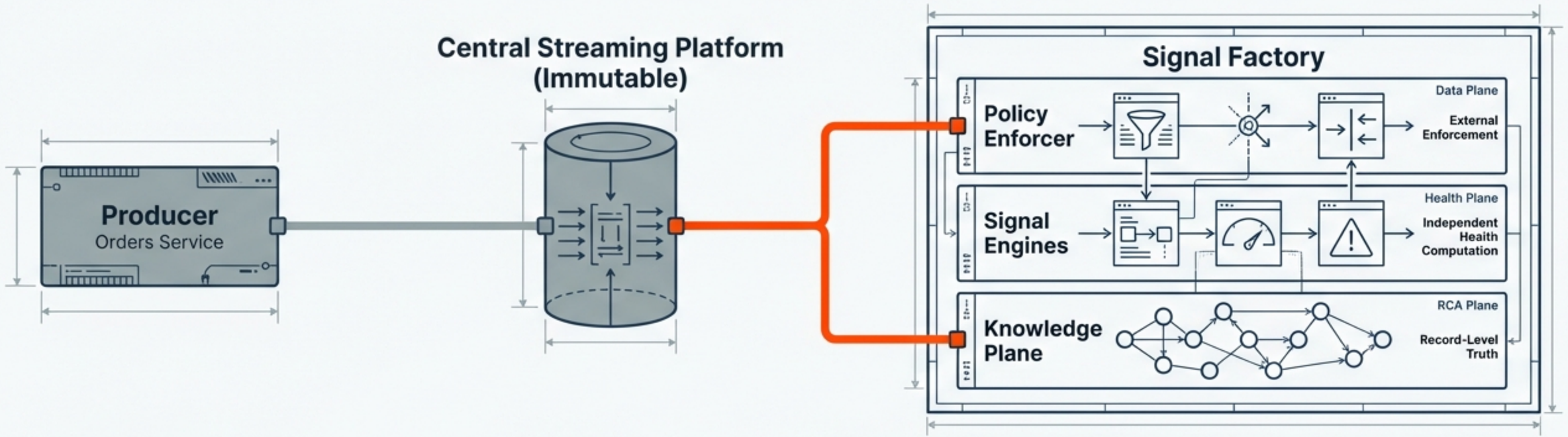
Achieve prevention-first data observability in a locked environment where the central streaming platform and producer code cannot be modified.

Scope

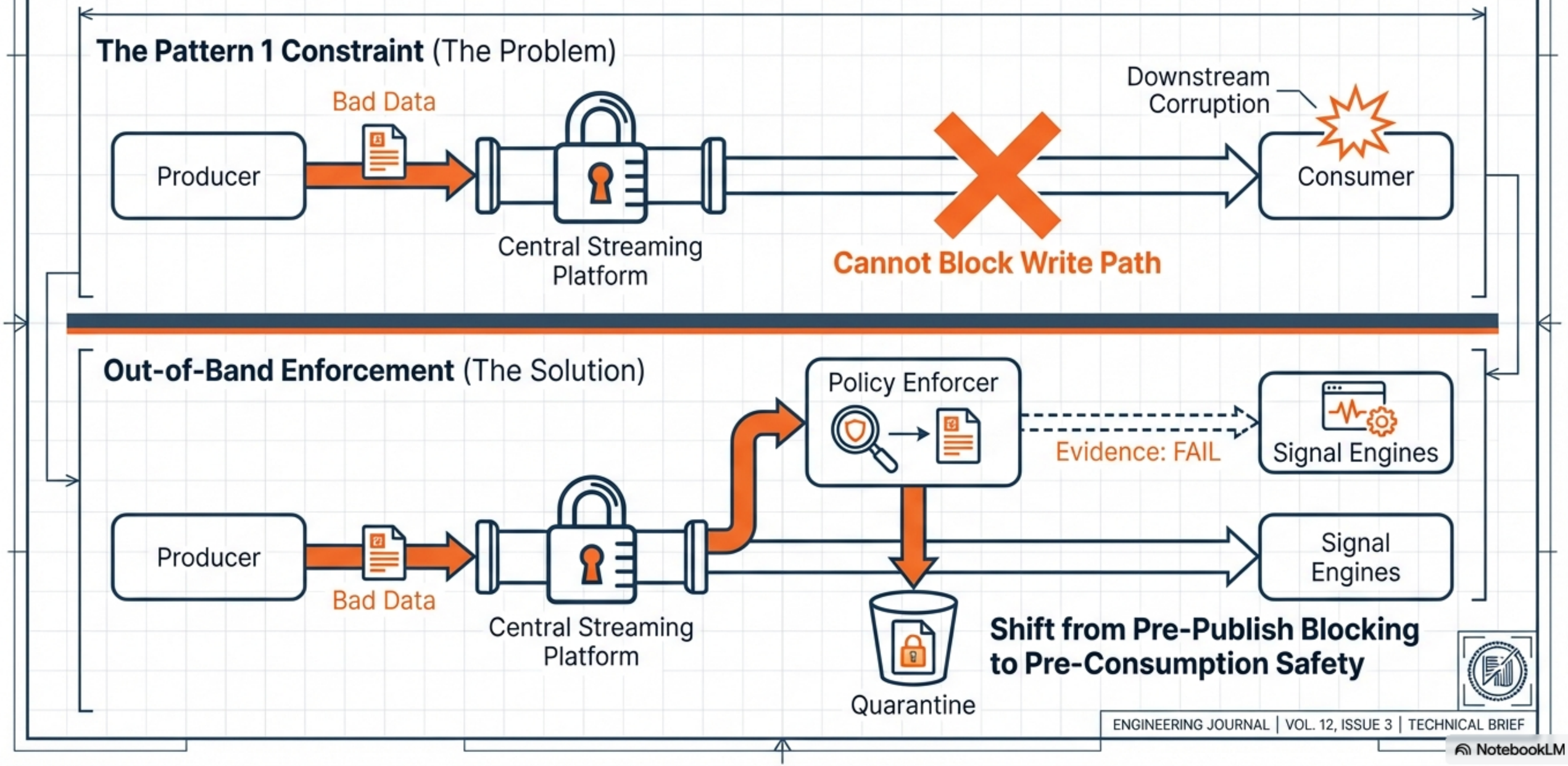
Detailed specification for the Policy Enforcer (Data Plane), Signal Engines (Health Plane), and the Knowledge Graph (RCA Plane).

Definition of Done

A system where record-level truth is established deterministically by enforcement, and system health is computed independently by signal engines.

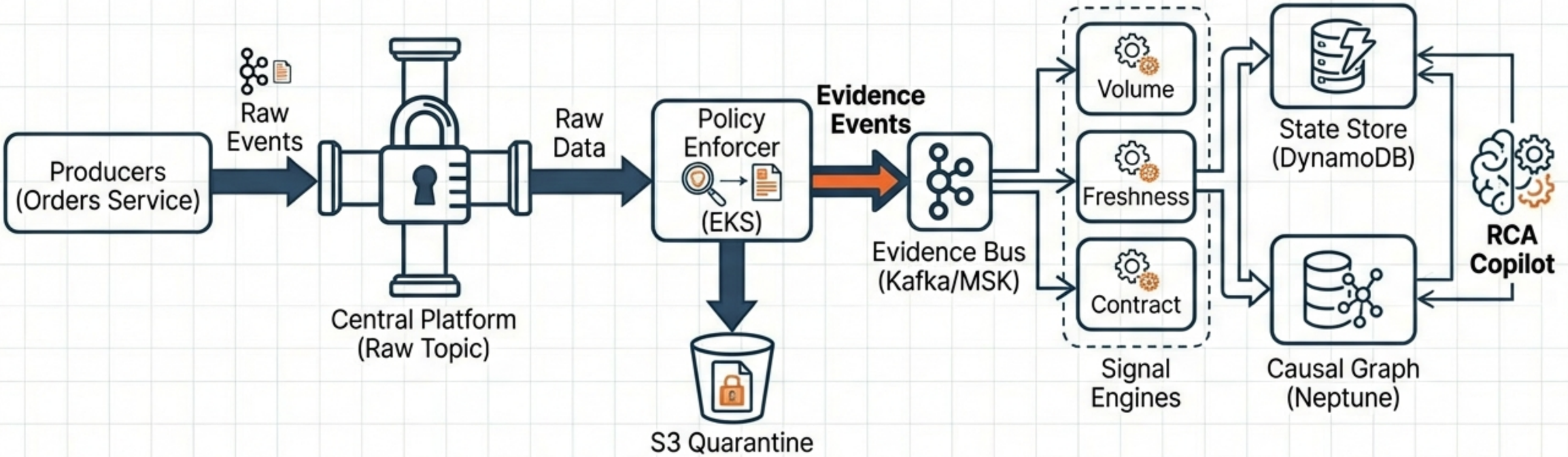


The Constraint: Immutable Producers and Platform



High-Level Data Flow

Raw Events Flow Once, Evidence Flows Everywhere



Architectural Note: Principle: The Policy Enforcer is the only component that touches raw data. Downstream engines operate exclusively on Evidence.

The Golden Rule: Separation of Responsibilities

Policy Enforcer (The Data Plane)

Record Truth

- **Question:** Is this specific record valid?
- **Action:** Parse & Canonicalize
- **Action:** Validate Schema & Contract
- **Action:** Detect PII
- **Output:** Immutable Evidence (PASS/FAIL)
- **Constraint:** NEVER computes rates, thresholds, or incidents.

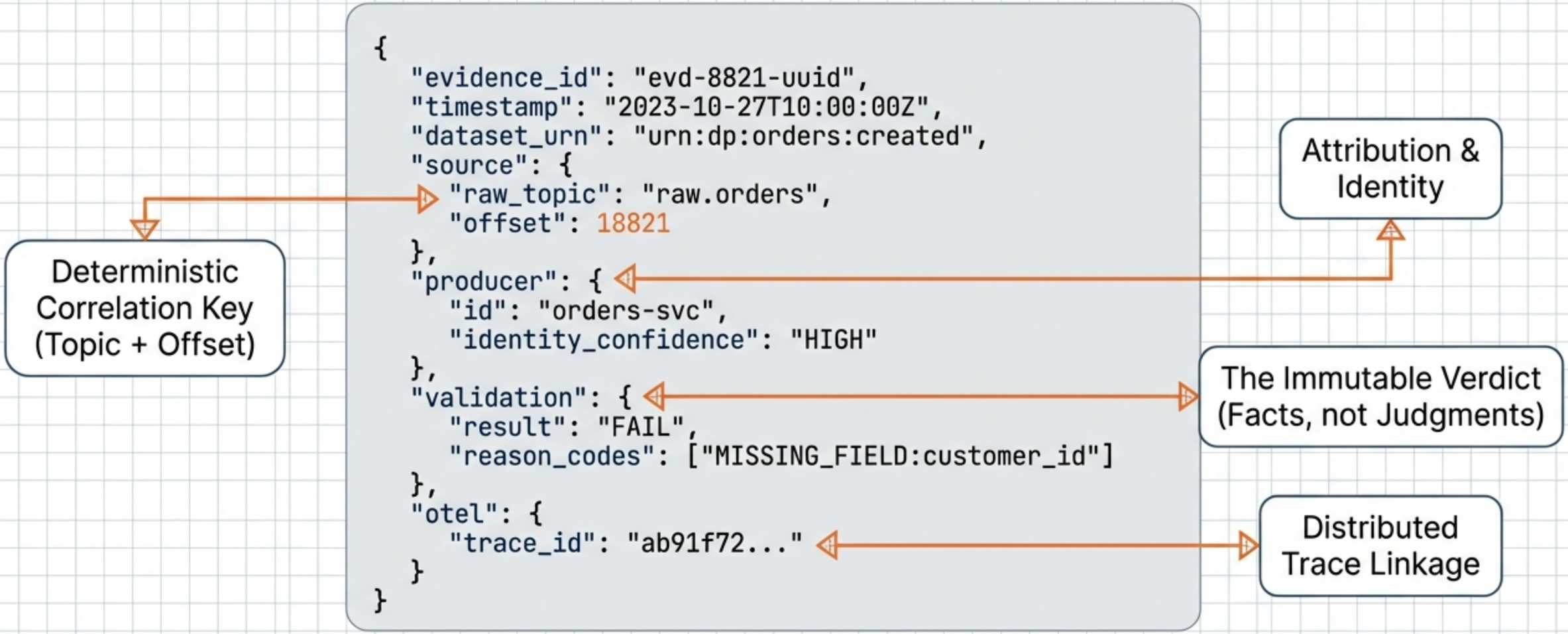
Signal Engines (The Health Plane)

System Health

- **Question:** Is the system healthy?
- **Action:** Aggregate Time Windows
- **Action:** Apply Thresholds & SLOs
- **Action:** Detect Anomalies
- **Output:** Signals & Incidents
- **Constraint:** NEVER re-parses payloads or re-validates schemas.

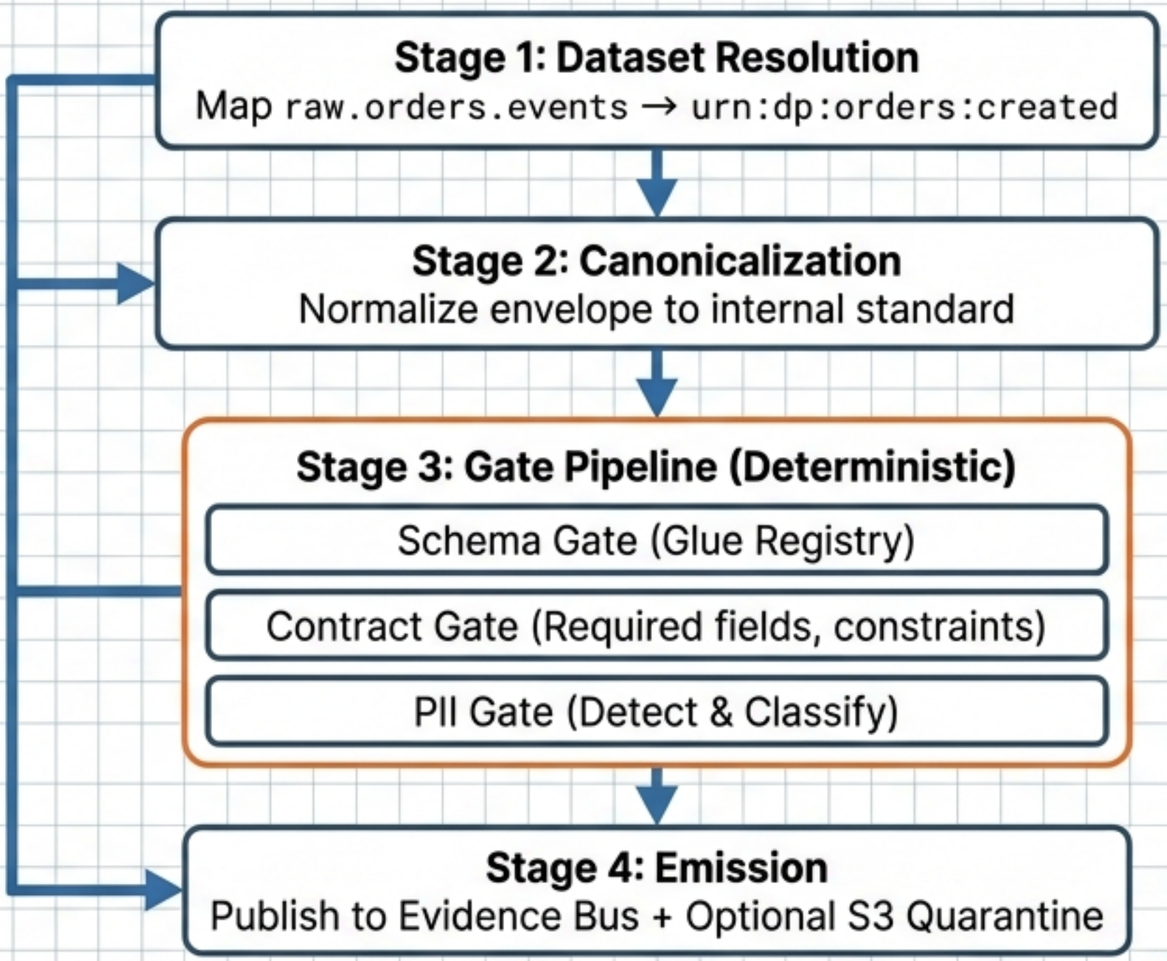
The Interface: The Evidence Event Schema

The **Immutable** API Between Layers



Component Deep Dive: The Policy Enforcer

The Runtime Data Plane Pipeline



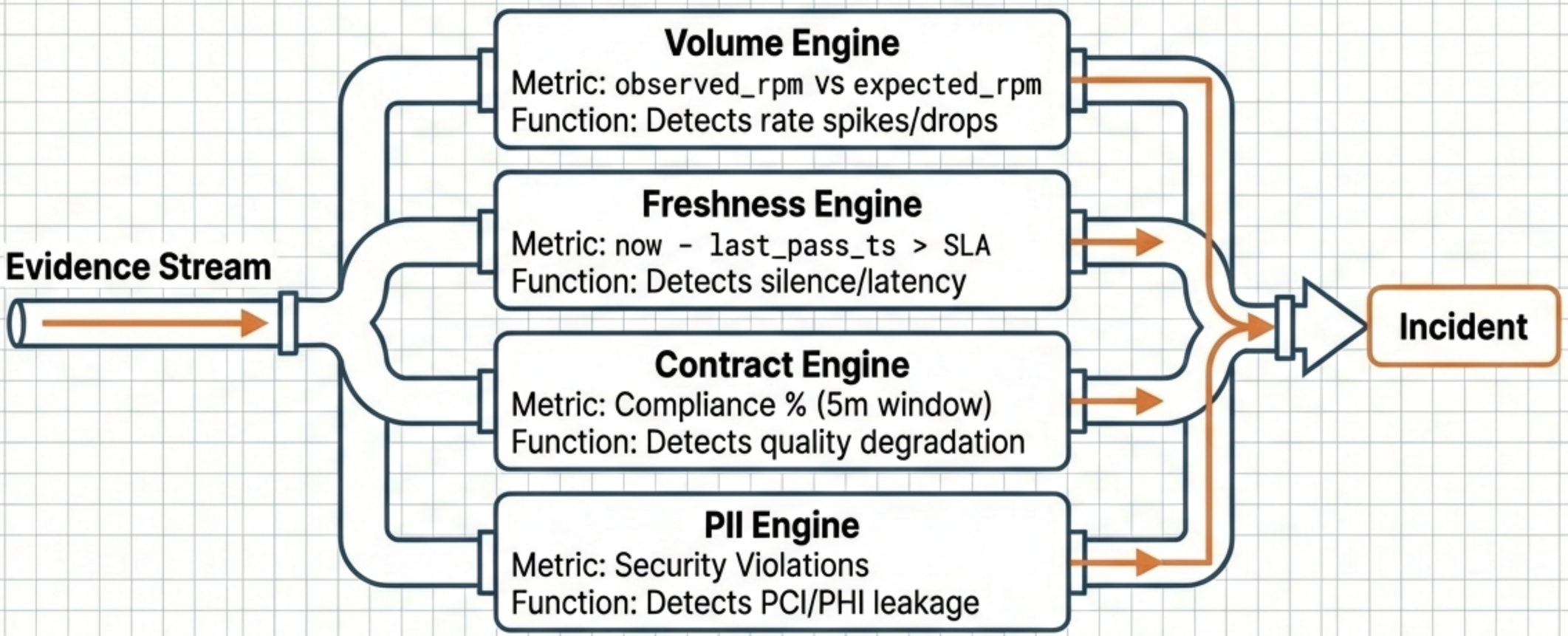
Performance SLOs

Evidence Completeness: ≥ 99.99%

Latency p95: ≤ 2s

Component Deep Dive: Signal Processing Engines

Computing System Health from Evidence



Output: Incidents containing links to evidence and traces, not just generic alerts.

The Steel Thread (Part 1): Ingestion & Enforcement

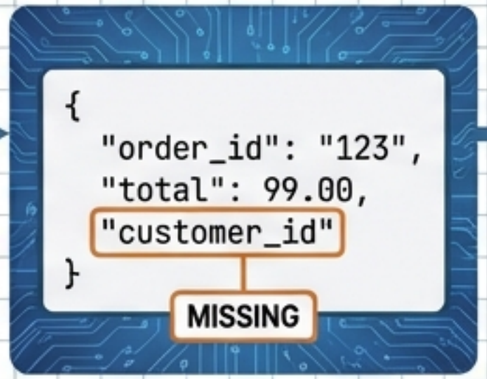
Scenario: The ‘Missing Customer ID’ Incident

09:58 AM

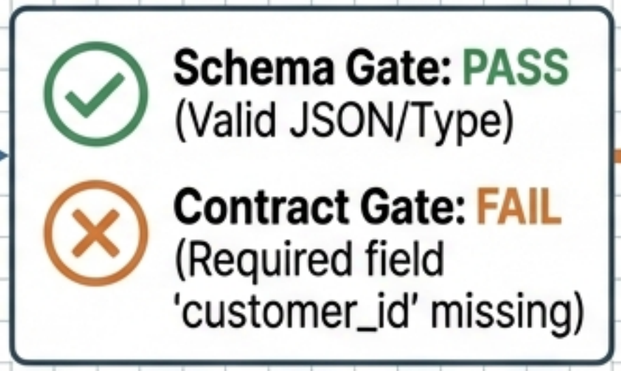
Orders Service Deploys Hotfix

10:00 AM

Raw Event Published



Policy Enforcer Gates

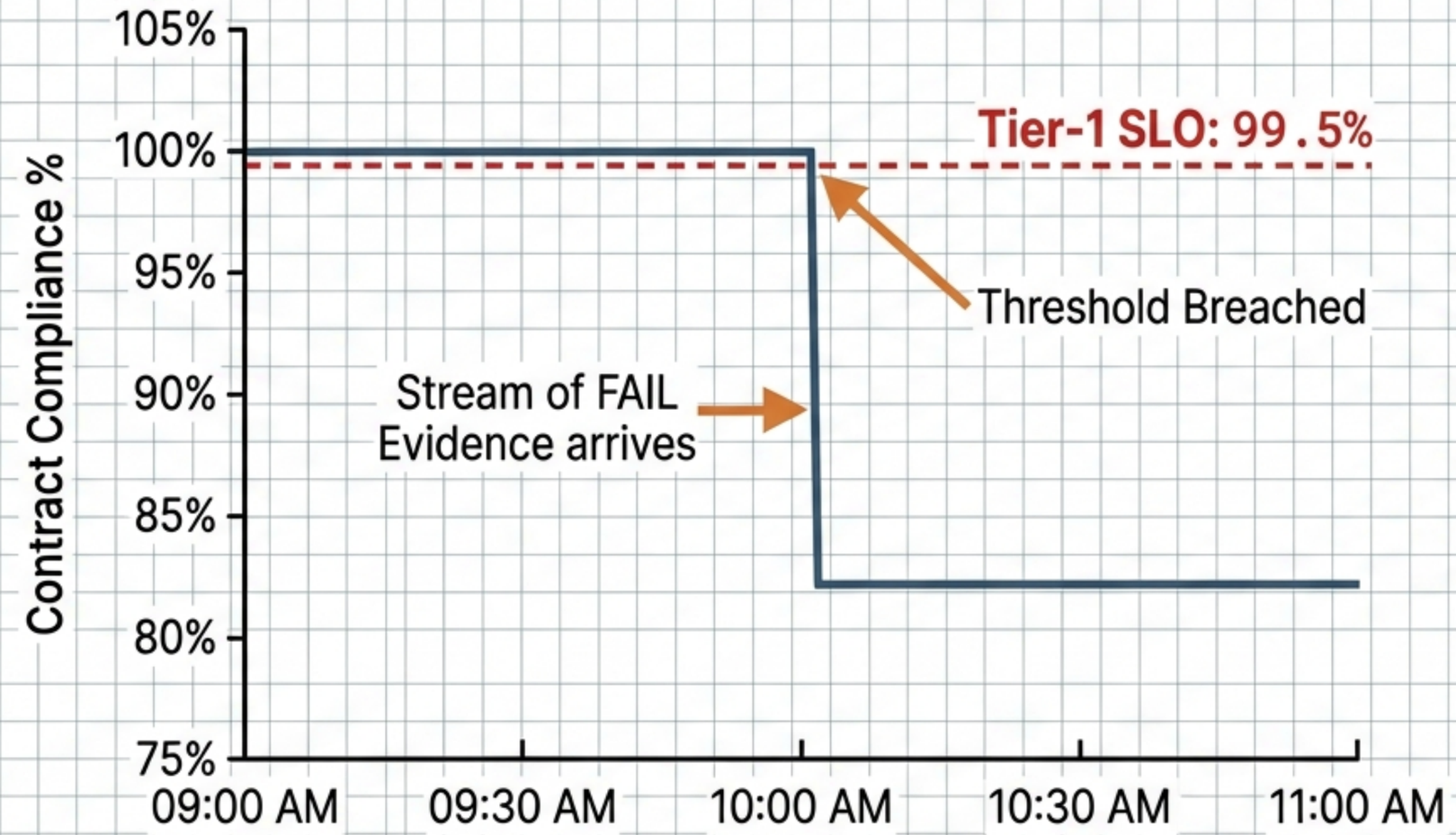


Outcome

Enforcer establishes the FACT of failure. It It does NOT page the on-call yet.

The Steel Thread (Part 2): Signaling & Incident

From Evidence to SEV-1



Incident Generated

SEV-1 Incident: Orders Service

Contract Violation > 0.5%
Linked Evidence: evd-8821

Operational State: DynamoDB Data Model

Answering “What is happening right now?”

DatasetRegistry (PK: dataset_id, SK: version)	
Purpose: Configuration	
`dataset_id, version`	SLOs, Contract IDs

SignalState (PK: signal_id, SK: timestamp)	
Purpose: Current Health	
Volume: Healthy, `Signal_id`	**Contract:** Critical

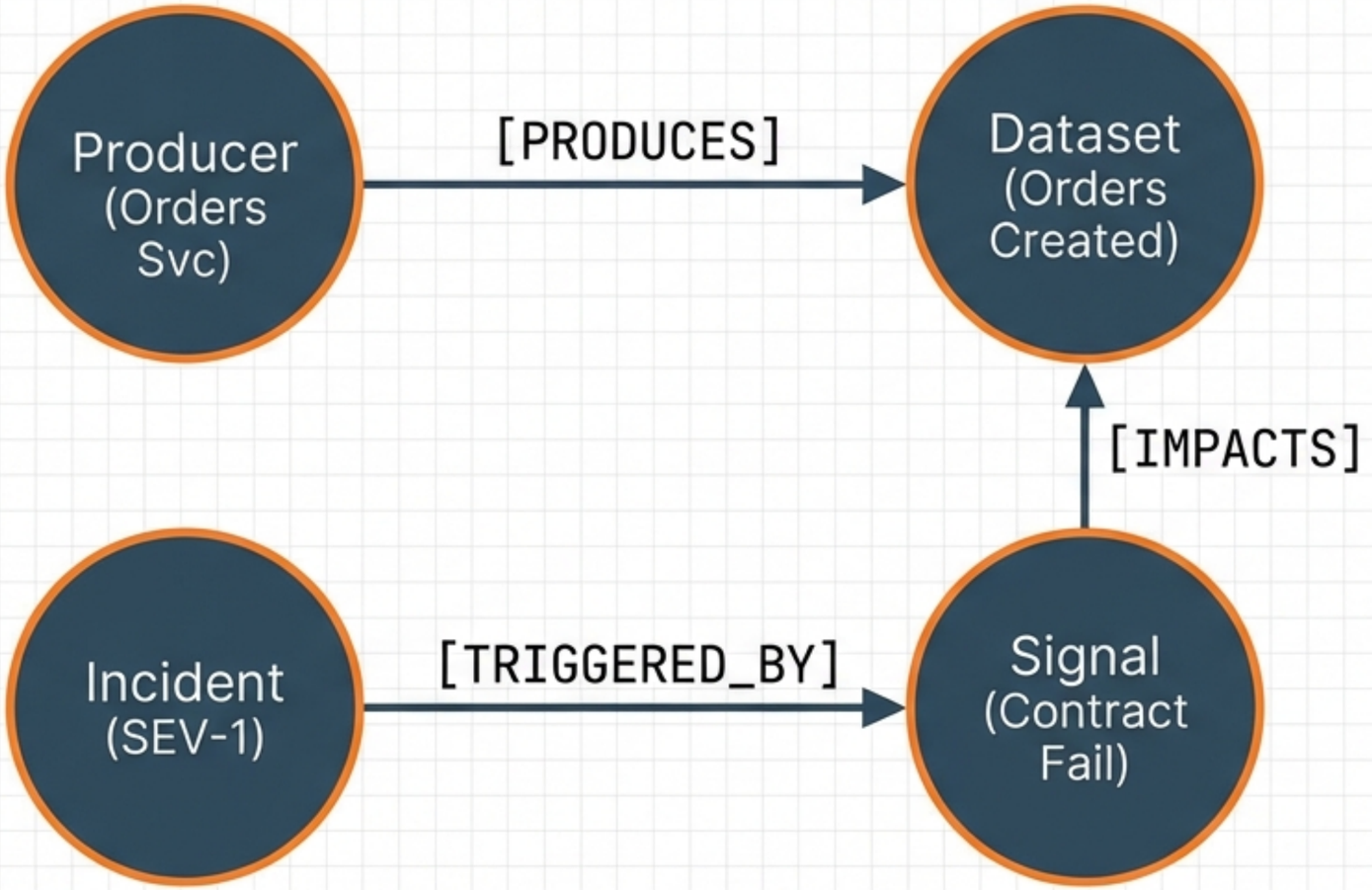
EvidenceIndex (PK: evidence_id, SK: timestamp)	
Purpose: Recent Lookups	
Data:	Recent FAIL evidence IDs (TTL: 7 days)
`evidence_id, Recent FAIL (TTL: 7 days)` `evidence_id, FAIL contrac (TTL: 7 days)`	

IncidentIndex (PK: incident_id, SK: timestamp)	
Purpose: Active Management	
Data:	Open Incidents
`Incident_id` `Incident_id`	... ...

Key Strategy: Optimized for fast point-lookups and dashboard backing.

Causal Graph: Neptune Data Model

Answering “Why did this happen?” (RCA)

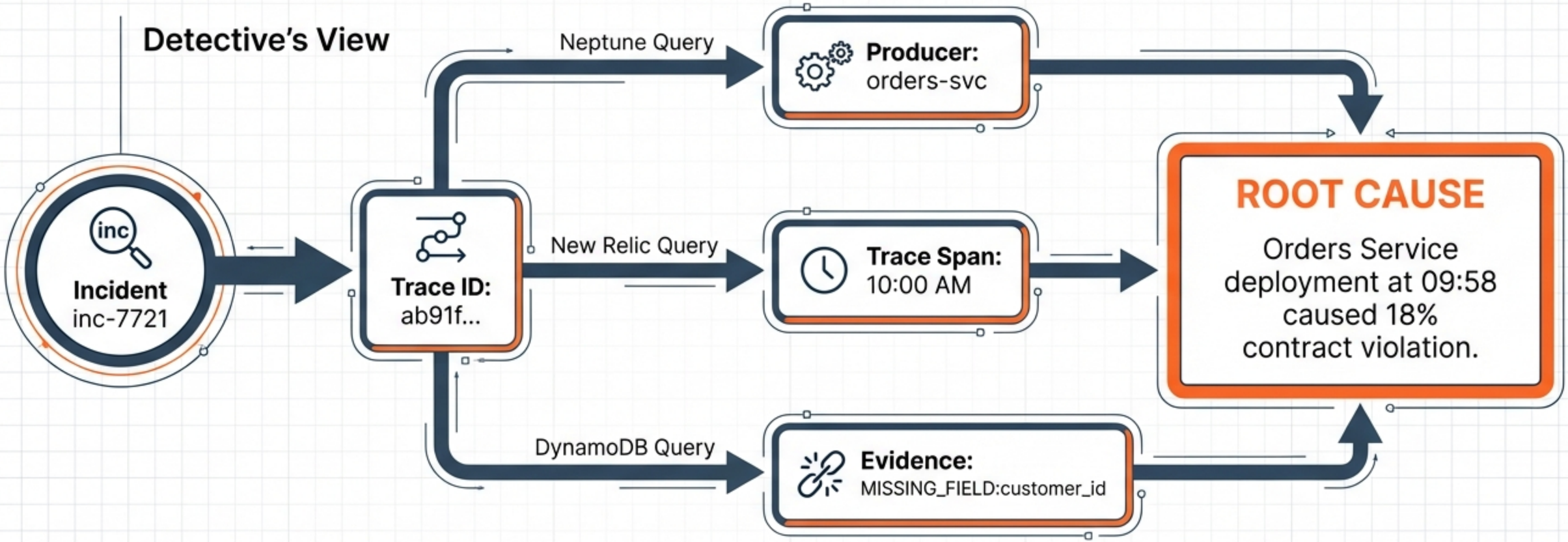


Cost Control

- What NOT to write: Every PASS evidence record, raw payloads.
- What TO write: Relationships, Incidents, and Failure Signatures.

Traceability & RCA Copilot

Deterministic Root Cause Analysis



Resilience: Failure Mode Analysis

Design for Failure

Scenario: Policy Enforcer Down		**Outcome:** Signal Engines detect “Evidence Gap”. Trigger `OBSERVABILITY_PIPELINE_DOWN` incident.
Scenario: Registry Unavailable		**Outcome:** Enforcer emits `'SCHEMA_REGISTRY_UNAVAILABLE'` evidence. Does NOT crash.
Scenario: Neptune Down		**Outcome:** Signals still fire. Incidents created via DynamoDB. Only RCA graph is delayed.

Implementation Rubric: Definition of Done

Engineering Validation Scorecard

Completeness

Evidence emitted for >99.99% of raw records.

Isolation

Signal computed within 60s of event arrival.

Latency

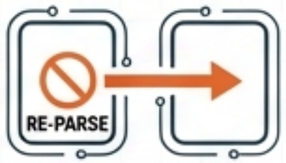
Signal computed within 60s of event arrival.

Traceability

Every incident provides a Trace ID or Deterministic Key.

Isolation

Signal engines NEVER re-parse raw payloads.

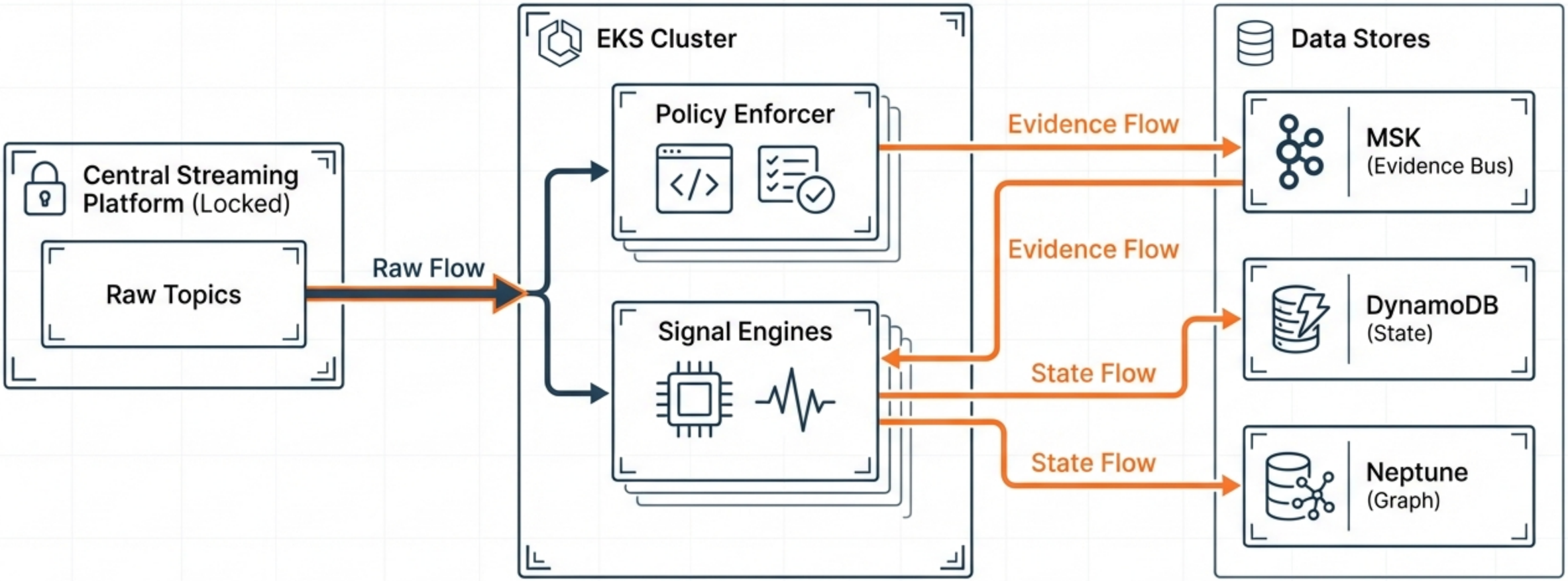


Cost Control

Zero ‘PASS’ evidence records written to Neptune.

Final Architecture: Pattern 1 Overview

Scale Without Fragility



Decoupling Observation from Judgment allows the platform to scale.