

H3-Claim-Agent // Technical Architecture Blueprint

Autonomous Forensic Claim Intelligence & Parity Defense on Google Cloud & Vertex AI

Target Environment: Google Cloud us-west2 (h3-healthcare-prod)	Active GCP Project: h3-claim-agent
Linked Billing Account: 012A95-92AF92-F722F1	Organization: highlandhealthhub.com
Version: v0.2.1-CloudNative (Production Specification)	Security Posture: Zero-PHI // Cloud DLP Fail-Closed // CMEK

1. Executive Overview & Core System Doctrine

Highland Health Hub (H3) operates the **H3-Claim-Agent**, an autonomous vertical AI SaaS platform built to neutralize systemic reimbursement shortfalls, algorithmic denials (CARC 197 / CARC 50), and federal Mental Health Parity and Addiction Equity Act (MHPAEA) violations in Behavioral Health & Substance Use Disorder (SUD) treatment.

The Core System Invariant: 'Give the model reasoning freedom inside a strictly bounded sandbox; govern all state, arithmetic, authority, and irreversible actions at the deterministic system boundary.' Large Language Models are strictly prohibited from calculating numbers, determining financial outcomes, or authoring unsourced appeal allegations.

2. The 7-Stage Google Cloud Native Pipeline

Stage	GCP Component	Operational Function & Deterministic Invariant
Stage 1: Intake	CMEK Cloud Storage gs://h3-claim-raw-intake	Custodial batch drops of 835/837P EDI, CSV remittances, and EHR exports secured by Cloud KMS (h3-storage-key).
Stage 2: Security	Cloud Sensitive Data Protection (Cloud DLP GATE-001)	Fail-closed Zero-PHI inspection. Scans SSN, DOB, name before any data enters reasoning tiers. Gate-002 verifies SHA-256 cross-matter isolation.
Stage 3: Reconciler	Deterministic Reconciler v2 (Pure Python Decimal)	Zero LLM Math. Enforces Underpayment = max(0, Allowed - Paid - PR - CARC 253) and CA Knox-Keene § 1371.35 15% statutory interest.
Stage 4: JEV Router	Cloud Tasks jev-shadow-queue	Pattern J3 triage routing administrative denials to Gemini 2.5 Flash and multi-issue clinical/parity disputes to Gemini 2.5 Pro.
Stage 5: Reasoner	Vertex AI Gemini 2.5 Pro / Flash	Sandboxed google-genai SDK execution with ADC, temp=0.0, Pydantic JSON schema binding, and citations strictly bounded to EV-xxx.
Stage 6: Admission	Two-Tier Audit Engine (audit.py)	Deterministic SHA-256 locator hash validation. Automatically quarantines hallucinated evidence citations with UNVERIFIED_EVIDENCE_CITATION.
Stage 7: Sinks	Firestore, BigQuery & Zero-PHI Neo4j Graph	Firestore operational state, BigQuery JEV shadow telemetry lake, and tokenized payer denial pattern mapping in Neo4j (:H3 namespace).

3. Dual Operating Modes: Facility SaaS vs. Litigation Support

Dimension	Mode A: Facility Revenue Recovery SaaS	Mode B: Legal Analytics-as-a-Service
Client Target	Behavioral Health & SUD Facilities (ASAM 1.0-3.7)	ERISA § 502 & Healthcare Litigation Law Firms
Intake Vault	gs://h3-vault-facility/<CLIENT_ID>/	gs://h3-vault-litigation/<MATTER_ID>/
Primary Deliverable	Executive Claim Intelligence Record (CIR), Revenue Opportunity Queue & Scrubber Rules	Master Litigation Damages Schedule, Statutory Interest Audit & Exhibit-Citable Dossiers
Legal Boundary	Administrative Provider Dispute Support	Attorney Work Product / UPL Protective Headers
Value Impact	Recovers working capital; keeps clinical beds open	Accelerates settlement; establishes trial damages

4. Benchmark Forensic Case Study: CASE-006 Financial Reconciler

To validate deterministic accuracy across the entire stack, **CASE-006** serves as our primary production proof point across live web telemetry, internal pytest validation, and whitepaper reporting:

Claim Attribute	Verified Value / Metric	Governing Legal / Technical Baseline
Case Identifier	CASE-006	Synthetic fixture canonical benchmark
Payer & Plan	Aetna Commercial PPO	Commercial PPO Out-of-Network Parity
Level of Care	ASAM 3.7 Inpatient Medically Monitored Detox	CPT H0010 // Severe Fentanyl protocol
Payer Denial Action	CARC 197 / RARC N702 (Precertification absent)	Unlawful step-down denial; retroactive auth valid
Billed Charge	\$2,500.00	Submitted UB-04 / 837I claim line item
Contract Allowed	\$1,800.00	Established commercial benchmark rate
Payer Paid Amount	\$1,200.00	Remitted via 835 ERA line adjustment
Patient Responsibility	\$0.00	Group Code PR isolated from carrier debt
Deterministic Underpayment	\$600.00	max(0.00, \$1800 - \$1200 - \$0 - \$0) via pure Decimal
Statutory Prompt-Pay Interest	\$44.38	CA Health & Safety Code § 1371.35 (15% p.a., 180-Day Lag)
Total Dollars at Risk	\$644.38	Direct recoverable carrier liability

5. Mathematical Infallibility Proof (180-Day Statutory Interest)

Daily Interest Factor: 0.15 div 365 equals 0.0004109589
Delinquency Span: 180 Calendar Days (Remittance Date minus Adjudication Deadline)
Statutory Interest: \$600.00 x (0.15 / 365) x 180 equals \$44.38356
Quantized Result (ROUND_HALF_UP): \$44.38

6. Architectural Governance & Readiness Gate

All 260 unit and integration tests across the H3 Claim Agent suite enforce these exact values. The public repository, live website, and API microservices maintain identical figures, guaranteeing zero mathematical variance between discovery exhibits, executive dossiers, and operational recovery queues.