Strategic Roadmap: Unifying Stem's AI-Driven Infrastructure

Target Role: Senior DevOps Architect **Primary Focus:** Consolidation of PowerTrack, Athena, and Locus Platforms **Timeline:** 90-Day Execution Sprint **Objective:** Dismantle architectural silos, establish a secure, standardized "Golden Path" platform, and drive significant cloud cost optimization through automated governance and self-service.

Executive Summary

Stem operates three distinct platforms—**Athena** (AI/ML workloads), **PowerTrack** (real-time asset monitoring), and **Locus** (location intelligence)—each with overlapping yet fragmented infrastructure. This "Three-Platform" model creates redundant services, security gaps, configuration drift, and inflated cloud spend.

This 90-day roadmap delivers a unified, **AI-optimized infrastructure foundation** via a **secure Golden Path** (opinionated best-practice patterns). Key outcomes include:

- **15–20% reduction** in operational cloud costs through consolidation and right-sizing
- **Full NIST/SOC 2-aligned continuous compliance**
- **30% faster time-to-market** for new features and models via self-service provisioning
- **Reduced MTTR** through centralized observability and tracing

By Day 90, the team will have a proven pilot migration, a bootstrap Internal Developer Platform (IDP), and a clear 12-month retirement path for legacy silos.

Phase 1: Discovery, FinOps & Immediate Risk Mitigation (Days 1–30)

Goal: Create a single source of truth for infrastructure, halt security and cost leakage, and build executive confidence with quick wins.

Key Initiatives

- **Comprehensive As-Is Audit & Service Inventory** Conduct automated discovery (using AWS Config, CloudQuery, or similar) across all three platforms. Map dependencies, redundancies (e.g., multiple RabbitMQ/Kafka clusters, duplicate Redis/Caching layers, overlapping S3 buckets), and underutilized AI resources (e.g., oversized GPU instances in Athena).
- **Advanced FinOps & Unit Economics Modeling** Implement granular AWS Cost Allocation Tags and Categories. Shift from account-level to **"Cost per AI Model / Service / Business Outcome"** attribution. Identify low-margin or zombie assets for immediate decommissioning.
- **Security & Compliance Baseline** Run automated scans (Checkov, tfsec, Trivy) against NIST 800-53 and SOC 2 controls. Prioritize critical findings (e.g., public EKS endpoints, unencrypted EBS volumes, missing IMDSv2).

Strategic Resolutions & Quick Wins

- **Risk:** Fragmented platforms expose public endpoints and lack centralized logging. **Resolution:** Enforce a **Global VPC Flow Logs** policy + AWS Security Hub aggregation. Create a prioritized remediation backlog for Phase 2.
- **Quick Win:** Disable/delete high-risk public resources and apply immediate tagging policies to capture ~5–10% cost savings in the first 30 days.

Phase 1 Deliverables

- **The "Efficiency Gap" Report** — Visual dashboard (e.g., via QuickSight or Grafana) showing spend breakdown, redundancies, and risk heat map.
- **Immediate Cost-Savings Action Plan** — List of 10–15 high-ROI actions with projected savings.

Success Metrics

- 100% visibility into cross-platform spend attribution
- Identification of ≥$X in annualized run-rate savings opportunities
- Zero critical vulnerabilities left unaddressed (P0 findings remediated or mitigated)

Phase 2: Design & Bootstrap the "Golden Path" Architecture (Days 31–60)

Goal: Establish Stem's standardized, secure, developer-friendly platform to eliminate toil and configuration drift.

Key Initiatives

- **AWS Control Tower + Organizational Guardrails** Deploy Landing Zone with multi-account strategy (e.g., separate accounts for dev/staging/prod, shared services, security logging). Enforce mandatory SCPs and guardrails.
- **Centralized IaC "Golden Path" Library** Build a mono-repo or curated Terraform Registry with reusable, versioned modules (VPC with private subnets, EKS with IRSA/OIDC, RDS/TimescaleDB clusters, S3 with encryption+KMS, etc.). Pin dependencies to Git commit SHAs.
- **AIOps & Unified Observability Stack** Consolidate monitoring into Datadog + CloudWatch (or equivalent), with unified dashboards, AI-driven anomaly detection, and cross-platform tracing (X-Ray + OpenTelemetry).

Strategic Resolutions

- **Risk:** Manual ad-hoc deployments lead to drift, pipeline failures, and supply-chain vulnerabilities. **Resolution:** Mandate **OIDC federation** for GitHub Actions workflows. Enforce immutable module sources (commit-hash pinning) and automated dependency scanning. Integrate DevSecOps gates (SAST/DAST) early.

Phase 2 Deliverables

- **Internal Developer Platform (IDP) Bootstrap** — Self-service catalog with pre-approved templates (e.g., "Deploy AI Inference Service" or "Provision Monitoring-Ready Asset Tracker").
- **Golden Path Documentation** — Including module catalog, usage guidelines, and security rationale.

Success Metrics

- End-to-end deployment of a NIST-compliant EKS cluster in **< 30 minutes** via self-service
- 100% of new infrastructure provisioned through Golden Path modules

Phase 3: Migration Execution, Velocity Acceleration & Team Enablement (Days 61–90)

Goal: Validate the platform through real workload migration and empower teams for long-term adoption.

Key Initiatives

- **Pilot Migration: Locus Non-Prod Service → Golden Path** Perform a "Lift and Reshape" (minimal changes + modernization) of a representative Locus service to the unified EKS cluster. Measure before/after metrics (cost, performance, reliability).
- **Unified CI/CD & DevSecOps Pipeline** Standardize on GitHub Actions with reusable workflows: build → test → SAST/DAST → deploy → post-deploy verification.
- **Developer Enablement Program** Launch **Architecture Decision Records (ADRs)**, platform "Office Hours", and training sessions. Roll out **Backstage** or Confluence-based Internal Developer Portal with "Click-to-Deploy" templates tailored for energy intelligence/AI workloads.

Strategic Resolutions

- **Risk:** Migration friction causes velocity drop and resistance. **Resolution:** Provide rich documentation, golden-path templates, and "migration accelerators" (e.g., automated refactor scripts, side-by-side environments).

Phase 3 Deliverables

- **Successful Production-Ready Pilot** — Migrated service running live with improved metrics.
- **12-Month Platform Retirement & Migration Roadmap** — Phased decommissioning of Athena/PowerTrack silos.

Success Metrics

- **50% reduction** in manual DevOps support tickets via self-service adoption
- Positive feedback from Athena and PowerTrack teams (e.g., via surveys)

Projected Multi-Year Business Impact

Impact Area	Year 1 Expected Outcome	Senior Architect Rationale
Operational Costs	15–20% Reduction	Elimination of redundant SaaS (Datadog, PagerDuty), right-sized AI/GPU instances, and automated shutdowns.
Security & Compliance	100% Continuous NIST/SOC 2 Compliance	Shift from point-in-time audits to always-on guardrails, automated scans, and policy-as-code.
Engineering Velocity	30% Faster Time-to-Market	Removal of infrastructure wait times; developers focus on models/features, not YAML battles.
Platform Reliability	Significantly Reduced MTTR	Unified observability + tracing enables rapid cross-platform root-cause analysis for asset failures.