

JAKE WILLIAMS

DATA & AI ENGINEER — DATA APPLICATIONS · SNOWFLAKE LAKEHOUSE ·
KAFKA CDC · AGENTIC AI (CLAUDE, MCP)

Denver, CO · info@jakeawilliams.com
linkedin.com/in/jakeintech ·
github.com/Jakeintech
jakeawilliams.com

Platform engineer at a specialty insurer: owner of the Kafka CDC streaming platform on Kubernetes, builder of core domains in the governed Snowflake lakehouse, and lead engineer of a production GenAI semantic layer. Ships data applications end to end — React/Next.js front ends on governed data — including two independent products in production. Spearheaded the enterprise **Claude plugin marketplace** (12 plugins, ~200 monthly users, analyst to C-suite) and cut \$30K+/yr in platform cost. Agentic analytics in production with MCP, model evaluations, and an audit trail that holds up under executive, auditor, and regulator scrutiny.

EXPERIENCE

Vantage Risk — Data Engineer III, Platform & Semantic Analytics

Mar 2024 — present

- **Spearheaded the enterprise Claude plugin marketplace** (12 plugins, 6 domains): ~200 monthly active users — 40% of the company, incl. 5 C-suite regulars — 2K+ queries/month; cut token consumption 30% vs. unassisted LLM use; own roadmap, evaluations, and governance.
- **Own the enterprise CDC/replication platform (real-time event streaming)**: inherited mid-build, shipped it, re-architected Event Hubs → Apache Kafka (Strimzi on Kubernetes), cutting **\$30K+/yr**; Debezium CDC from SQL Server, PostgreSQL, and Oracle (LogMiner), plus custom Java Kafka Connect connectors where none existed.
- **Lead engineer, semantic analytics** (Claude + Snowflake Cortex): governed NL analytics with consistent regulatory reporting — Quarterly Reserve Review **3 days** → **<30 min**, QBR prep **4 days** → **<1 hr**; Cortex evaluations made answers trusted and compliant.
- Analytics engineering at scale: **700+ dbt models** (ELT) across 5 data domains on a Snowflake medallion lakehouse, incl. a financial-reconciliation framework; type-2 period-fact models for point-in-time analysis; data-quality tests in CI.
- Dagster data orchestration — partitioned software-defined assets, jobs, schedules, asset checks; run-failure sensors alert fast; business-driven sensors on bound deals improved closure **20%**.
- GitOps: Argo CD + Helm deliver every platform service — Kafka, connectors, orchestration, observability; PySpark compaction jobs on Apache Iceberg keep streaming tables query-efficient.
- Observability as code: **60+ dashboards/monitors** ship via GitHub Actions (Datadog, Prometheus); forecast monitors prevented 5 OOM incidents and avoided ~20% infra cost (FinOps).
- Leadership: own pull-request review workflows; build GitHub Actions pipelines of all kinds — build/test/lint matrices, deployment, release automation, dashboards-as-code; mentoring, office hours driving AI-assisted development (Claude Code) adoption; SOC 2 SOPs, audit support.

Vantage Risk — Reporting Analyst & Architect

Mar 2023 — Mar 2024

- Co-designed enterprise data architecture (lake, EDW, reporting) on the design committee; helped build the reporting solution the company scaled on. **Promoted to Data Engineer in 12 months.**
- Built a key component of the enterprise pricing tool — a TypeScript microservice on Azure integrating GraphQL, MCP, and PostgreSQL APIs; SME on feasibility, cost, and risk.

State Farm — Data Analyst

Jan 2021 — Mar 2023

- Drove data governance for specialty P&C analytics — integrity, lineage, regulatory compliance; delivered an Agile migration of analytics to AWS Redshift; designed and ran team AWS training.

Independent Contractor — data, analytics & technology

2016 — 2021

- Built client data, analytics, and technology platforms — flagship: Laughlin River Tours' full stack (automated pipelines, AWS QuickSight over Glue); Stripe payments and menu systems for a casino venue; a table-management app with **\$100K+ cost avoidance**.

INDEPENDENT PRODUCTS — IN PRODUCTION

- **303Happenings** (303happenings.com) — content-driven events & city-data platform for Denver (block-editor CMS), designed, built and operated solo: Next.js/React/TypeScript + PostgreSQL; multi-source ingestion with entity resolution; 14 city-data maps; Event API, MCP server, pgvector semantic search; multi-agent LLM runtime (Claude, model failover) with tool use, LLM-as-judge review, eval-measured learning loop and memory, permission-gated actions, audit ledger.
- **BarkHouse** — booking/CRM platform for pet-care businesses (Next.js/React/TypeScript, Stripe). (barkhouseapp.com)

CORE ENGINEERING

SQL · Python · PySpark ·
TypeScript/JavaScript · React ·
Next.js · HTML/CSS · UI/UX · REST
APIs · JSON · Terraform · Go · Java
(Kafka Connect) · distributed
systems · CI/CD · GitHub Actions ·
VS Code · Agile

STREAMING & PLATFORM

Apache Kafka · Kafka Connect ·
Debezium · CDC · Kubernetes ·
Docker · Strimzi · Helm · ArgoCD ·
GitOps · AWS · Azure

LAKEHOUSE & ANALYTICS

Snowflake · dbt · Dagster · Apache
Iceberg · PostgreSQL · medallion
architecture · dimensional
modeling · ELT · data quality in CI

AI & AGENTIC SYSTEMS

Claude · MCP · AI agents · agent
orchestration · RAG · LLM
evaluation · guardrails · semantic
layers · Snowflake Cortex
· LLMOps

GOVERNANCE & OPS

Data governance · lineage · data
quality · SOC 2 · FinOps · Datadog
· Prometheus · observability
as code

CERTIFICATIONS

AWS Certified Solutions
Architect — Associate
AWS Certified Cloud Practitioner

EDUCATION

B.S. Business Administration
& Management, Northern
Arizona University

INDEPENDENT WORK

Waybill — open-source Claude
Code plugin: deterministic token
accounting for AI-assisted work,
MIT
(github.com/Jakeintech/waybill)
QuotedAI — iOS app on the App
Store (SwiftUI, HealthKit; on-
device LLM via Apple Foundation
Models; 2K+ downloads)