

KALYAN KASINADHUNI

🏠 205 N Harvard St, Vermillion, SD (Open to Relocation)
✉ kasinadhunikalyan@gmail.com — 📞 +1 (605) 202-9653
in LinkedIn — 📁 Portfolio

PROFESSIONAL SUMMARY

Data Engineer with 4+ years designing enterprise-grade data solutions for Fortune 500 clients in healthcare (Bayer) and fintech (Go Audits). Expert in building petabyte-scale data platforms with real-time streaming (Kafka/Flink), cloud optimization (AWS/Azure), and MLOps pipelines. Delivered \$3M+ in cost savings, 5x performance improvements, and AI-driven automation for critical business processes. Certified in AWS, Azure, and Databricks with hands-on GenAI implementation experience.

PROFESSIONAL EXPERIENCE

Go Audits, Delaware, USA

Aug 2023 – May 2025

Data Engineer

- Designed and deployed a distributed event processing system using Kafka Streams with Flink Stateful Functions, handling 500K transactions/sec across 3 global regions with exactly-once processing semantics. Reduced fraudulent transaction slippage by 37% (\$4.2M annual savings).
- Spearheaded AWS FinOps initiative implementing Redshift RA3 instances with automatic workload management, S3 Intelligent Tiering with lifecycle policies, and Glue Flex jobs with spot instances. Achieved 48% cost reduction (\$320K annual savings) while maintaining 99.95% SLA.
- Engineered an LLM-powered document processing system using Llama2-70B, LangChain, and Pinecone, automating extraction of key terms from 10K+ monthly contracts. Reduced legal review time by 62% while maintaining 98% accuracy via human-in-the-loop validation.
- Architected a multi-tenant Delta Lake on EMR with Unity Catalog governance, implementing MERGE operations for SCD Type 2 dimensions and Z-ordering optimization for 100+ concurrent analysts. Improved complex join performance by 8x compared to legacy Parquet.
- Built a time-travel capable feature store using Feast on Redis, enabling reproducible training sets for fraud detection models. Implemented point-in-time correctness for 75+ features across 2-year lookback windows.
- Developed Terraform modules for enterprise AWS deployment (VPC peering, Transit Gateway, PrivateLink) with GitHub Actions CI/CD enforcing policy-as-code (Checkov, Terrascan). Onboarded 12 new teams in 3 months with zero security exceptions.
- Deployed OpenLineage with Marquez across Spark/Airflow pipelines, tracking 500+ datasets with column-level lineage. Enabled impact analysis for GDPR compliance with 1-hour traceability (down from 3 days manual effort).
- Re-engineered Glue PySpark jobs using Ray on EMR with automatic scaling (5-200 workers), reducing runtime for daily financial reconciliations from 4.5 hours → 22 minutes (92% cost reduction).
- Established center of excellence for data engineering, mentoring 6 engineers on Spark tuning (broadcast joins, skew handling) and cost-aware development. Reduced average job costs by 65% across team projects.
- Designed attribute-based access control (ABAC) for Snowflake using Okta integration and dynamic data masking. Achieved SOC2 Type 2 certification with zero critical findings during audit.

Data Engineer I

- Led migration of 22 legacy SAS pipelines to Delta Lake on Databricks, implementing incremental MERGE patterns for 12TB/day of patient telemetry. Reduced FDA submission preparation time from 14 days → 72 hours.
- Built Kafka Connect pipeline with Debezium CDC, streaming 800+ Oracle tables to Synapse Real-Time Analytics. Enabled sub-5-second latency for critical care device monitoring across 50 hospitals.
- Implemented Synapse workload management with materialized views and result-set caching, reducing monthly spend by \$22K while improving query performance 3x.
- Developed Great Expectations suite with 500+ validation rules (statistical, semantic, regulatory), increasing data trust scores from 78% → 99.4%. Automated alerts via PagerDuty integration reduced issue resolution time by 85%.
- Productionized PyTorch clinical trial models using MLflow Model Registry with A/B testing endpoints. Achieved 50ms inference latency for patient risk scoring at 2K RPS.
- Converted 150 SSIS packages to Azure Data Factory using parameterized ARM templates and Git version control. Implemented data contract testing via dbt for cross-team dependencies.
- Applied Delta Lake Z-ordering and optimized file sizes (128MB) to achieve 15-minute SLA for critical patient cohort analytics (previously 4 hours).
- Collaborated with Bayer Germany to implement OMOP CDM standard across US/EU datasets, enabling cross-border research studies with 98% mapping accuracy.
- Deployed Azure Purview with custom scanners for 120+ sensitive data sources, achieving automated PII/PHI classification with 95% recall.
- Prototyped natural language to SQL using LangChain + OpenAI, reducing analyst query development time by 40%. Published internal whitepaper on GenAI for clinical data exploration.

TECHNICAL SKILLS

Cloud Platforms	AWS (Kinesis, Glue, Redshift, Lambda, EMR), Azure (Databricks, Synapse, Purview), GCP (BigQuery, Dataflow)
Big Data Stack	Spark (PySpark/Scala), Kafka, Flink, Airflow, dbt, Snowflake, Iceberg, Delta Lake, Hudi
Data Architecture	Data Mesh, Medallion Architecture, CDC (Debezium), DWH Modeling (Data Vault 2.0)
ML/GenAI	LLM Orchestration (LangChain, LlamaIndex), Vector DBs (Pinecone), Feature Stores (Feast)
DevOps	Terraform, Kubernetes, CI/CD (GitHub Actions, Jenkins), Docker, ArgoCD

EDUCATION & CERTIFICATIONS

M.S. Computer Science University of South Dakota (GPA: 3.5/4.0)

2023–2025