Dileep Kumar Reddy Kapu

Senior Data Engineer — Cloud Engineer — Big Data Engineer
Albuquerque, NM — Open to Relocation
dileepkreddy5@gmail.com — +1 (505) 364-5197 — linkedin/dileep-kumar-reddy — github/dileepkreddy5

Professional Summary

Senior Cloud/Data Engineer with 6+ years of experience designing and leading enterprise-scale data platforms in HIPAA-regulated healthcare environments. Expert in AWS, Azure, Snowflake, and distributed Spark architectures, with a track record of modernizing legacy systems, establishing governed lakehouse foundations. Focused on building secure, scalable, and cost-optimized platforms that drive measurable business impact.

Key Skills

- **Cloud & Lakehouse Platforms:** AWS (S3, Glue, EMR, Athena, Lambda, IAM, KMS, CloudWatch), Azure (Fabric, ADF, Synapse), Snowflake, Delta Lake (Medallion), Apache Iceberg
- **Data Engineering Architecture:** ETL/ELT Pipeline Design, dbt (Analytics Engineering), Spark, PySpark, Airflow, Kafka, Batch & Streaming Processing, Schema Evolution, Dimensional Modeling, Data Quality Validation
- **Infrastructure, Security & DevOps:** Terraform (Infrastructure as Code), CI/CD (GitHub Actions), Docker, Kubernetes, RBAC, Secure Networking, Observability, Performance Tuning
- **Programming & Data Systems:** Python, SQL, Spark SQL, MySQL, REST API Integrations, Linux
- **GenAI & Retrieval Systems:** RAG Architectures, LangChain Workflows, Embeddings, Vector Databases (Pinecone, Milvus), Chunking Strategies, Prompt Context Engineering, Knowledge Graph Integration

Professional Experience

New Mexico Health Care Authority — Santa Fe, NM
Senior Data Engineer — Jun 2025 – Present

- Led implementation of secure multi-cloud ingestion architecture bridging Azure clinical systems with AWS analytics, establishing a HIPAA-compliant platform serving 1M+ residents with 99.9% reliability.
- Standardized Terraform-based infrastructure patterns across networking, IAM, and data services to eliminate configuration drift, enforce security baselines, and enable repeatable, auditable multi-environment deployments.
- Designed a Snowflake-centric ELT architecture leveraging AWS Glue and Snowpipe to consolidate 20M+ fragmented health records, reducing query latency by 40% and enabling near real-time statewide analytics.
- Built CI/CD pipelines using GitHub Actions and Terraform to replace manual deployments, enabling zero-downtime schema evolution while improving release reliability and deployment velocity.
- Implemented platform-wide security controls using IAM, VPC isolation, KMS encryption, and CloudWatch observability to enforce least-privilege access, reducing audit exceptions by 35% and improving incident MTTR by 25%.

Optum (Client: UnitedHealth Group) — Albuquerque, NM
Senior Data Engineer (Contract) — Jun 2024 – May 2025

- Architected and governed 25+ production-grade pipelines integrating 12+ disparate healthcare data domains, processing 3–5TB monthly to enable scalable enterprise analytics across claims, member, and provider ecosystems.
- Built Delta Lake–based ELT frameworks to address schema drift and incremental processing challenges, enabling reliable analytics consumption across 7+ downstream claims, member, and provider reporting teams.
- Orchestrated batch workflows using Airflow with idempotent DAGs and automated backfills to eliminate manual reruns, reducing operational intervention by 40% and improving data consistency across recurring workloads.
- Re-engineered Spark workloads through partition strategy redesign, caching optimization, and resource allocation tuning, reducing processing latency by 35% and accelerating new data source onboarding by 25%.

University of New Mexico (UNM Health — Information Technologies) — Albuquerque, NM
Data Engineer — Jan 2023 – May 2024

- Built a serverless ingestion pipeline using AWS SES, Lambda, S3, and Python-based validation logic to replace manual registration processing, reducing 200K+ annual submissions from hours to under five minutes while

improving data quality controls.

- Designed and maintained a 40+ table MySQL data warehouse using structured ETL workflows and optimized SQL transformations, centralizing academic and administrative datasets for 500+ stakeholders and improving reporting accuracy and consistency.
- Automated registration and payment workflows using Lambda, API Gateway, and Python-driven REST integrations to eliminate 70% of manual effort, enabling reliable execution of 250+ conference sessions during a five-day event.

Carelon Global Solutions (Elevance Health) Bangalore, India
Big Data Engineer Jun 2020 – Dec 2022

- Engineered large-scale batch and streaming pipelines using Spark, Hive, and Airflow to process healthcare provider and claims datasets, improving data availability and reliability for enterprise analytics teams.
- Led the migration of enterprise analytics workloads from on-prem Hadoop to AWS-managed services, redesigning processing architectures to improve platform reliability and increase operational efficiency by 15% through elastic infrastructure adoption.
- Automated repetitive DataOps workflows using Python-based frameworks, generating **$250K+** in annual operational savings.

Leadership & Impact

- Acted as multi-cloud SME at NMHCA governing secure, HIPAA-compliant data architecture decisions.
- Led and mentored 4+ engineers delivering resilient healthcare analytics platforms at scale.
- Mentored 10+ engineers building revenue-generating data platform driving $120K annually.
- Received High Impact Award for automation initiatives saving $250K+ operational costs.

Certifications

- AWS Certified Solutions Architect – Professional
- Microsoft Certified: Azure Solutions Architect Expert (AZ-305)
- AWS Certified Data Engineer – Associate
- AWS Certified Machine Learning – Specialty
- HashiCorp Certified: Terraform Associate (004)

Projects

Iceberg Lakehouse with Terraform *(AWS S3, Glue, Athena, Apache Iceberg, Terraform)*

- Architected a serverless Iceberg-based lakehouse on S3 to enable ACID transactions, schema evolution, and time-travel analytics while provisioning reproducible infrastructure using Terraform.

Real-Time Streaming Data Platform *(Kafka, AWS MSK, Lambda, Spark Structured Streaming)*

- Designed an event-driven streaming architecture using Kafka and Spark Structured Streaming to process high-velocity data with low-latency guarantees and scalable enrichment pipelines.

Education

University of New Mexico, Albuquerque, NM Master of Science in Computer Engineering - Dec 2024
Coursework: Distributed Systems, Advanced Cloud Computing Architectures, Database & Query Optimization

Gitam University, India Bachelor of Engineering in Electronics & Communication Engineering - May 2021