

Roshan Dhamala

Senior Data & AI Engineer | Data Platform Architecture

Dallas, TX | 313.455.8561 | drdhamala@gmail.com | linkedin.com/in/rdhamala

PROFESSIONAL SUMMARY

Senior Data Engineer with 5+ years of experience designing and scaling enterprise **data and AI platforms** using Databricks, Apache Spark, PySpark, Python, SQL, dbt, and AWS. Experienced in lakehouse architecture, distributed processing, enterprise migrations, data modeling, batch and streaming systems, governance, reliability, performance optimization, and **Generative AI** solutions including LLMs, RAG, vector search, document processing, and AI agents.

EXPERIENCE

Veeva Systems

Pleasanton, CA

Senior Data Engineer

Jan. 2025 – Present

- Served as **technical architect for Veeva's enterprise data platform**, scaling it from an initial POC to a production Databricks lakehouse (**100+ TB, 120+ datasets**) and defining ingestion, modeling, and governance standards adopted across enterprise engineering and analytics teams.
- Owned the **data architecture for the Salesforce-to-Veeva Vault CRM migration**, designing reconciliation and schema-mapping patterns that migrated millions of records with **zero downstream analytics disruption or reporting data loss**.
- Own **end-to-end architecture** for batch and near-real-time enterprise integrations across SaaS, REST APIs, AWS Kinesis, and SFTP, establishing the **medallion (bronze/silver/gold) reference design** used by downstream teams.
- Re-architected high-cost Databricks workloads using Photon, incremental processing, partition pruning, and Spark tuning, cutting **compute spend by 37%** with no SLA regression.
- Established the platform's **governance and reliability framework** spanning Unity Catalog, data contracts, lineage, RBAC, and automated data quality, supporting **50+ downstream consumers**.
- Architected **LLM and RAG workflows** for document processing, structured extraction, metadata enrichment, vector search, and enterprise search, enabling agentic workflows over governed enterprise data.
- Lead architecture reviews and mentor engineers on Spark optimization, data modeling, and production design patterns while partnering with analytics and business teams on technical solution design and delivery.

Data Engineer

Jan. 2022 – Jan. 2025

- Built and operated production **batch and streaming data pipelines** using Databricks, Apache Spark, Delta Lake, AWS Kinesis, REST APIs, and enterprise systems, reducing processing latency by **60%**.
- Standardized orchestration, CI/CD, testing, monitoring, and alerting using Apache Airflow, Databricks Jobs, and Git while developing scalable **data models and ETL/ELT pipelines** with PySpark, SQL, and dbt.

Texas Tech University Health Sciences Center

Lubbock, TX

Software Developer Intern

Jan. 2021 – Dec. 2021

- Designed Snowflake data models and automated incremental ingestion using Apache Airflow and Snowflake Tasks, improving query performance by **45%** and enabling near-real-time reporting.

Discover Financial Services

Riverwoods, IL

Software Engineering Intern

Jun. 2021 – Aug. 2021

- Developed ETL workflows for large-scale financial transaction data, including profiling, validation, and anomaly detection to support SOX-compliant analytics.

TECHNICAL SKILLS

Languages & Processing: Python, PySpark, SQL, Scala, Bash, Apache Spark, Spark SQL, Structured Streaming

Data Platforms & Engineering: Databricks, Delta Lake, Snowflake, Apache Iceberg, Apache Kafka, dbt, Apache Airflow, Delta Live Tables, Data Modeling, ETL/ELT, Batch & Streaming, REST APIs, Tableau

Cloud & Infrastructure: AWS (S3, Kinesis, Glue, Lambda, EMR), Azure, GCP, Terraform (IaC), Docker, Kubernetes

DevOps, Governance & Reliability: Git, GitHub Actions, CI/CD, Unity Catalog, Data Contracts, Schema Evolution, Data Lineage, RBAC, Great Expectations, Data Quality, Observability, Monitoring, SLA Management

AI & Generative AI: LLMs, RAG, AI Agents, Agentic Workflows, Embeddings, Vector Search, Document Processing, Structured Data Extraction, Prompt Engineering

CERTIFICATIONS

Databricks Certified Data Engineer Associate | AWS Certified Cloud Practitioner

EDUCATION

Texas Tech University, Lubbock, TX | M.S. Software Engineering, Dec. 2021 | B.S. Computer Science, Dec. 2020