

SHAMANTH REDDY ROMPICHERLA

Data Engineer | Snowflake • Databricks • PySpark • AWS / Azure
US | +1 (217) 758-5010 | Shanthompicherla@gmail.com | LinkedIn

SUMMARY

Data Engineer with 4 years building cloud-native ELT pipelines and Lakehouse platforms across healthcare and insurance. Migrated 50TB+ of legacy data to Snowflake and cut multi-hour batch windows to under three hours, processing 100M+ records daily on PySpark, Airflow, and DBT. Strong in dimensional modeling, Kafka streaming, and Terraform-driven CI/CD, with hands-on experience extending pipelines for LLM / RAG retrieval use cases.

SKILLS

Languages: Python, SQL, PySpark, Scala, Shell Scripting

Data Warehousing & Lakehouse: Snowflake, Databricks, Delta Lake, Apache Iceberg, dbt, Dimensional Modeling (Star Schema, SCD Type-2)

Big Data & Streaming: Apache Spark, Kafka, Airflow, Hadoop, Hive

AWS: S3, Glue, Lambda, EMR, Step Functions, Athena, IAM, CloudWatch, Redshift

Azure: Data Factory (ADF), ADLS Gen2, Synapse, Functions, Key Vault, DevOps

Modern Data / AI: LLM Data Pipelines, RAG, Vector Databases (Pinecone, Weaviate), Embedding Pipelines, LangChain

Databases: Snowflake, Oracle, SQL Server, PostgreSQL, MongoDB, DynamoDB

DevOps & IaC: Docker, Kubernetes, Terraform, Jenkins, GitHub Actions, Git, CI/CD

Data Quality & Governance: dbt tests, Great Expectations, RBAC, Data Lineage, Data Masking, HIPAA Controls

BI & Visualization: Power BI, Tableau, Looker

EXPERIENCE

Data Engineer | UnitedHealth Group | Dallas, TX.

Feb 2026 – Present

- Built production ELT pipelines on Snowflake and PySpark integrating claims and provider data, reducing a key clinical analytics window from roughly 8 hours to under 3.
- Tuned Snowflake warehouses through clustering keys, micro-partition pruning, and right-sized compute, lowering query cost on the highest-spend workloads by an estimated 30%.
- Extended a document-ingestion pipeline with vector embeddings and LangChain to support an internal LLM-based clinical search prototype used by analytics teams.
- Automated deployments with Terraform and GitHub Actions, standardizing dev, staging, and production environments and shrinking release cycles.
- Implemented HIPAA-aligned access controls (RBAC, dynamic data masking) across patient and claims datasets, passing internal security review with no critical findings.

Data Engineer | MetLife | Chicago, IL

Aug 2024 – Jan 2026

- Re-engineered legacy ETL into PySpark distributed jobs across underwriting and claims, roughly halving daily batch runtime and removing the primary pipeline bottleneck.
- Delivered Azure Data Factory, Snowflake, and Databricks pipelines supporting analytics over 25M+ insurance policies, cutting reporting turnaround for downstream teams.
- Authored dbt transformation models using star-schema dimensional design with SCD Type-2 handling, plus dbt and Great Expectations tests that raised data-quality pass rates across claims analytics.
- Built Kafka streaming for fraud and claims event monitoring, moving event processing from batch latency to near-real-time and enabling proactive fraud flags.
- Automated cloud infrastructure with Terraform and Azure DevOps CI/CD, reducing manual provisioning effort across multiple data engineering teams.
- Engineered feature pipelines for predictive underwriting and retention models, accelerating data-science model development cycles.

Data Analyst | Accenture | Hyderabad, India

Aug 2021 – Jul 2023

- Analyzed 50TB+ of legacy Oracle, SQL Server, and Hadoop data migrated to Snowflake for Fortune 500 clients, identifying reporting gaps and improving query performance by an estimated 60%.
- Built and maintained SQL-based reporting and reconciliation frameworks across AWS ELT pipelines (Glue, S3, Airflow), supporting batch reporting cycles compressed from roughly 12 hours to under 3.
- Performed row-level data validation and reconciliation during CDC cutover from legacy systems to Snowflake, ensuring data accuracy and parity for downstream analytics.
- Designed reusable data profiling and validation templates that cut new-source onboarding effort by an estimated 40%, speeding up analysis turnaround across migration programs.
- Analyzed Snowflake query and partitioning performance to identify cost drivers, contributing to an estimated 30% reduction in cloud processing spend.
- Automated reconciliation and validation reporting in Python and SQL, replacing manual verification across millions of migrated records.

EDUCATION

Master of Science, Computer Technology — Eastern Illinois University 2024

Bachelor of Engineering, Computer Science — REVA University 2021