

VASANTHI ROUTHU

Austin, Texas, USA | vasanthilatha.bi@gmail.com | 949-424-2466

[LinkedIn](#) | [GitHub](#)

PROFESSIONAL SUMMARY

Senior Data Engineer with 7+ years of experience architecting high-scale ETL/ELT solutions in highly regulated industries. Expert in Python, SQL, and Azure/Snowflake/AWS cloud ecosystems. Proven track record of managing petabyte-scale data lakehouses while ensuring strict adherence to Data Governance, Security, and Compliance (HIPAA/PHI). Skilled at optimizing cloud compute costs and translating complex healthcare requirements into performant, reliable data architectures.

TECHNICAL SKILLS

Data Architecture: Medallion Lakehouse (Bronze/Silver/Gold), Delta Lake, Data Lakes, Real-time Streaming, Data Security, HIPAA Compliance, PII De-identification

Data Modeling: Star Schema, Snowflake Schema, Dimensional Modeling, Data Vault 2.0, Semantic Modeling

Programming: Python, SQL (T-SQL, PostgreSQL, PL/pgSQL), PySpark, DAX

Big Data & Frameworks: Databricks, Apache Spark, Delta Tables, Airflow, Kafka, Skyflow Privacy Vault

Cloud Platforms: Microsoft Azure (ADF, Databricks, Synapse), AWS (S3, Redshift, Glue), Snowflake

Databases: Snowflake, Amazon Redshift, PostgreSQL, Azure SQL, SQL Server, Cosmos DB

Performance & Cost: Query Tuning, Table Partitioning, Clustering, Batch Processing, Cloud Cost Optimization

Methodologies: GitHub, GitFlow, CI/CD (GitHub Actions, Azure DevOps, GitLab), DataOps, Agile/Scrum

PROFESSIONAL EXPERIENCE

Senior Data Engineer

GoodRx, USA June 2024 – Present

- Architected a config-driven 3-layer **Medallion pipeline in Databricks** to process 40-plus attributes from JSON sources; engineered deduplication logic and automated Slack/PagerDuty alerting for real-time failure monitoring.
- Orchestrated end-to-end data delivery via **Airflow**, automating S3 unloads to client SFTPs and utilizing AWS Glue to synchronize transaction data into Amazon Redshift.
- Engineered a **HIPAA-compliant de-identification engine** for more than 100 million historical records; integrated Skyflow for PII detokenization and orchestrated coordinated, multi-sink updates across S3, Redshift, and PostgreSQL.
- Optimized **PostgreSQL performance**, reducing runtime from **14 hours to less than an hour** by implementing batch-level commits and conditional trigger logic to eliminate table locks on near real-time data.
- Designed enterprise-grade dimensional models and data marts that enabled self-service analytics for more than 200 internal stakeholders and product owners.
- Developed a real-time streaming integration using **Kafka and Airflow** to ingest, transform, and load live telemetry into Amazon Redshift; maintained high data freshness for mission-critical dashboards.

Senior Data Engineer

Cision, Hyderabad, India August 2021 – August 2022

- Led the migration of legacy data architectures to **Snowflake** using Azure Data Factory (ADF) and Spark; implemented RBAC and dynamic data masking to strengthen governance standards.
- Engineered **dynamic schema handling processes** within Azure-based pipelines to automate table structure updates, significantly reducing manual maintenance and enhancing ETL flexibility.
- Optimized large-scale SQL transformations, improving end-to-end data availability by **40 percent** for ad-hoc research; designed Power BI dashboards to provide stakeholders with real-time data health visibility.
- Championed Engineering Excellence by providing technical training and documentation for junior engineers; drove performance tuning initiatives that reduced processing bottlenecks across the Azure suite.

Data Engineer

Itron, Hyderabad, India *December 2017 – July 2021*

- Architected and Migrated an end-to-end BI ecosystem from on-premises to **Azure**; successfully transitioned legacy SSRS/SSIS packages to ADF and moved SSAS cubes to Azure Analysis Services.
- Modernized manual taxation processes by developing a robust automated ETL framework, replacing Excel-based workflows with centralized data pipelines and **Power BI** financial dashboards.
- Engineered a custom **RPA solution using Python and Selenium** to automate the generation, renaming, and multi-sink distribution of parameterized paginated reports.
- Developed high-performance SSAS Tabular models and complex DAX semantic layers; optimized T-SQL stored procedures to support high-concurrency financial reconciliation and global audits.
- Managed the full deployment lifecycle of reporting suites, ensuring 100 percent data accuracy and consistent delivery for critical financial reconciliation windows.

EDUCATION

Master of Science in Computer Science — University of North Texas, USA (2024)

Bachelor of Technology in Electronics and Communication Engineering — JNTU, India (2018)

CERTIFICATIONS

Microsoft Certified: Azure Data Engineer Associate (DP-200/DP-201)

Microsoft 70-761: Querying Data with Transact-SQL

Microsoft 70-778: Analyzing and Visualizing Data with Power BI

ADDITIONAL INFORMATION

No sponsorship required

Based in Austin, TX; Open to 100% Remote or Relocation within the USA