

VASANTHI ROUTHU

[Austin, Texas, USA](#) | vasanthilatha.bi@gmail.com | [\(949\)424-2466](#) | [LinkedIn](#) | [Portfolio](#) | [GitHub](#)

SUMMARY

Sr. Data Engineer with 7+ years of experience building secure, scalable, and high-performing data pipelines across AWS, Azure, and GCP. Expert in Python, Apache Spark, Databricks, Airflow, and Redshift, with hands-on leadership in architecting end-to-end data platforms including semantic layers for real-time and batch analytics. Proven ability to mentor junior engineers, ensure operational stability, and deliver data-driven solutions that support business intelligence, compliance, and capital markets workflows. Adept at collaborating across Agile teams, applying DevOps practices, and leveraging cloud-native services like Glue, S3, Lambda, Step Functions, and Azure Purview.

EXPERIENCE

Sr Data Engineer, GoodRX

February 2024 - Present, Remote, US

- Developed and maintained critical ETL pipelines using Apache Airflow, PySpark, and Databricks, processing over 10+ TB/day, achieving 99.9% uptime.
- Automated ingestion from SFTP + Kafka to AWS S3, reducing manual work by 90% and failure rate to <4%.
- Migrated 20+ pipelines from GitLab CI/CD to Airflow, redesigned storage architecture from BigQuery to Amazon Redshift, reducing latency by 60%.
- Designed Delta Lakehouse architecture on Databricks for petabyte-scale data with Parquet, Spark SQL, and semantic layers, doubling query performance and cutting compute costs by 20%.
- Built real-time streaming analytics with Kafka + Spark, enabling <15s latency dashboards for decision support.
- Enforced CCPA & HIPAA compliance through automated PII/PHI de-identification pipelines and deletion workflows.
- Mentored junior engineers, conducted critical code reviews, and led focused Agile standups to proactively resolve blockers.
- Championed observability by implementing monitoring/logging pipelines using custom Airflow alerts and Slack integrations.
- Collaborated with data scientists and ML engineers to integrate machine learning models into production data pipelines.
- Implemented risk mitigation strategies for data quality, security, and compliance across cloud platforms.

Sr. Data Engineer, Cision

August 2021 - August 2022, Hyderabad, India

- Designed and implemented end-to-end ETL/ELT pipelines using Azure Data Factory (ADF) and Azure Databricks to process large-scale financial and capital markets data from transactional systems, market feeds, and regulatory reports.
- Developed reusable metadata-driven frameworks to dynamically create tables and ingest datasets, enabling scalable schema evolution and reducing manual intervention across pipelines.
- Built an automated data integration process using ADF, streamlining ingestion from multiple sources and saving over 20 hours per week in manual processing.
- Configured ADF orchestration workflows to manage complex, multi-step data pipelines, achieving a 40% reduction in processing time.
- Collaborated with business stakeholders to gather requirements and translate them into actionable Power BI reports, reducing manual reporting efforts by 98.9%.
- Engineered real-time agent performance dashboards in Power BI, integrating with Azure Synapse Analytics for cross-regional aggregation and insights.
- Implemented role-based access controls (RBAC) using Azure Active Directory (AAD) to enforce data security for claims adjusters and underwriters.
- Ensured regulatory compliance by applying data governance practices using Azure Purview, Key Vault, and RBAC for sensitive customer data.
- Worked in Agile squads with PMs, analysts, and QA engineers to deploy weekly release.
- Applied change management practices to guide data platform adoption and process improvements.

Data Engineer, Itron

December 2017 - July 2021, Hyderabad, India

- Automated the taxation process by designing a BI workflow that replaced manual Excel consolidation, aggregation, and analysis—reducing processing time by 80%, enhancing efficiency, and improving accuracy for the tax team.
- Implemented robust SSIS error handling, logging, and security systems, enabling real-time error detection and resolution that improved data quality and minimized downtime.
- Enhanced decision-making by integrating Paginated Reports in Power BI with Row-Level Security, securing access for 75+ users and ensuring 100% compliance with data privacy policies.
- Successfully migrated 30+ SSIS packages to Azure Data Factory (ADF), improving pipeline scalability by 60% and reducing maintenance overhead by 25%.
- Developed Power BI and Paginated Reports to automate financial reconciliation, reducing manual effort by 80% and increasing data accuracy by 30% through streamlined aggregation and validation.
- Collaborated with cross-functional teams to align data solutions with business objectives and strategies.
- Supported capital markets workflows by delivering timely, accurate financial and regulatory data reports.

EDUCATION

Master of Science in Computer Science

University of North Texas • Texas, USA • 2024 • 4

Bachelor of Technology in Electronics and Communication Engineering

Jawaharlal Nehru Institution of Technology • Hyderabad, India • 2018

SKILLS

- **Databases:** Redshift, Postgres, SQL Server, Azure SQL Database, Oracle 11g, Cosmos DB, Snowflake (SnowSQL)
- **Cloud Platforms:** AWS (S3, Redshift, Glue, Lambda), Azure Data Factory (ADF), Azure Databricks, Azure Synapse Analytics, Azure Analysis Services (AAS)
- **Big Data Technologies:** Apache Spark, Hadoop, Databricks (SparkSQL, Delta Lake, PySpark, etc), Azure Data Lake Storage (ADLS), Azure Data Explorer (ADX)
- **BI & Reporting Tools:** Power BI, Tableau, SSIS, SSAS, SSRS, SSDT, PowerPivot, Power View, Excel
- **Programming Languages:** T-SQL, Python, R, PowerShell
- **Version Control & DevOps:** Git, Azure DevOps, CI/CD
- **Containerization & Orchestration:** Docker, Kubernetes, Azure Kubernetes Service (AKS), Databricks CLI, Containerized ETL Workflows
- **Project Management & Methodologies:** Agile, Scrum, Waterfall, Jira

CERTIFICATIONS

- DP-200 Implementing Azure Data Solutions
- DP-201 Designing an Azure Data Solutions
- 70-761 Querying Data with Transact-SQL
- 70-778 Analyzing and Visualizing Data with Microsoft Power BI.