Harish Kumar Sarathi

Data Engineer

7575 Frankford Rd, APT 1923, Dallas, TX

+1 (945) 216 4304 | sharishkmr@gmail.com | linkedin.com/in/shk30 | github.com/Harish1230s

Eligible to work in the US for internships and full-time for up to 36 months without sponsorship

Summary

Aspiring Data Engineer with experience building scalable ELT data pipelines and analytics dashboards using Airflow, Spark, Azure, and AWS. Skilled in SQL, Python, and cloud-native tools with a focus on distributed data systems. Currently pursuing MS in Business Analytics & AI, passionate about applying data science to energy systems and sustainability.

Education

University of Texas at Dallas, Texas   *December 2025*
Master of Science, Business Analytics and Artificial Intelligence (GPA: 3.63 / 4.00)

Vellore Institute of Technology, Vellore, India   *May 2022*
Bachelor of Technology, Electronics and Communication Engineering - IoT and Sensors (GPA: 3.08 / 4.00)

Professional Experience

A Consultancy Services, Houston, Texas   *May 2025 – Present*
Business Analyst Intern - Data Science & Cloud

- Developed scalable ELT pipelines using Azure and Apache Airflow; automated ingestion for structured logs and optimized Spark-based workflows for dashboarding.
- Provided L1/L2 Azure support, provisioning diagnostic pipelines and supporting job orchestration with Airflow and Git version control.
- Collaborated using MS 365 tools including SharePoint and OneDrive for documentation and cross-functional team alignment.

Karthik Roller Flour Mills Pvt. Ltd., Vellore, India   *August 2022 – July 2024*
Supply Chain Analyst

- Automated procurement and reporting workflows using Tableau and Excel macros, reducing weekly reporting time by 40%. Resolved 50+ support tickets and documented root causes to prevent recurrence.
- Conducted feasibility studies for vendor consolidation and built KPI dashboards used by procurement and finance teams, improving communication, inventory planning accuracy and reducing raw material overstock by 20%.
- Documented 3+ system upgrade cycles across procurement software, enhanced automated data flows, and worked with IT to deploy versioned reports for compliance, improving audit readiness by 30%.

Delta Electronics India Pvt. Ltd., Krishnagiri, India   *Mar 2022 – Jul 2022*
Operations Engineering Intern

- Conducted 150+ thermal/electrical VFD tests & improved QA by 12% through root cause & failure trend analysis.
- Debugged test failures with QA & revised workflows, reducing process bottlenecks & improving throughput by 19%.
- Managed build-to-packaging process for sales orders, ensuring compliance with specs and timely customer delivery.

Skills & Certifications

Languages & Scripting: Python, SQL, Scala (basic), Bash, PowerShell, Java, C++, HTML
Big Data & Workflow: Apache Spark, Airflow, Hadoop (HDFS, Hive), Redshift, Glue
Cloud Platforms: Azure (ADF, VMs, RBAC), AWS (EC2, S3, Lambda, RDS, IAM), GCP (BigQuery, Composer)
Analytics & Visualization: Tableau, Power BI, StatsModels, Excel, Adobe Analytics
DevOps & Tools: Git, GitHub, Docker, Kubernetes, Jira, Microsoft 365
Concepts: ETL/ELT, Distributed Systems, Data Warehousing, Query Optimization, Dashboards, UAT, CI/CD
Certifications: AWS Certified Cloud Practitioner (June 2025); Snowflake SnowPro Core Certification (May 2025)

Projects

Modern Data Platform – Banco Wild West (OBAP) | AWS, Glue, Redshift, SageMaker

- Built zoned S3 lakehouse using Glue ETL and Spark jobs; version-controlled pipelines using Git and deployed ML inference via SageMaker endpoints.
- Deployed containerized inference endpoints via Docker and Kubernetes, enabling scalable ML prediction services.

Customer Behavior Modeling | Python, Poisson/NBD, StatsModels , Machine Learning Models

- Forecasted repeat purchases using Poisson/NBD and zero-inflated models for targeted business marketing.
- Evaluated model performance via AIC/BIC and BI dashboards to guide campaign design optimization.

Achievements

Dean's Excellence Scholarship, University of Texas at Dallas
Merit-based academic award received at admission for MS in Business Analytics & Artificial Intelligence.

DataWhiz 2025 – 2nd Place, EnVision UTD (Team Innovate Y), Safety Excellence Group
Analyzed OSHA injury data using LDA topic modeling; built logistic regression model to accurately predict hospitalization risk. Delivered an interactive Power BI dashboard that guided strategic decision-making in workplace safety.