

Data Engineer with 2 years of experience building scalable batch and streaming pipelines across GCP, Azure, and on-prem environments using Python, Java, Spark, and Airflow. Skilled in metadata-driven ETL, dimensional modeling (SCD, CDC), and data quality frameworks. Experienced in Docker, Kubernetes, and CI/CD automation (Azure DevOps, Cloud Build) for HIPAA-compliant, cloud-native deployments.

SKILLS

Languages & Scripting	Python, Java, SQL, Bash, C, C++, JavaScript, TypeScript
Cloud Platforms	Google Cloud (BigQuery, GCS, Pub/Sub, Composer, Dataflow), AWS (S3, Lambda, Redshift, EMR), Azure (DevOps, Storage, Functions)
Data Engineering Stack	Apache Spark, Apache Kafka, Airflow, Apache Beam, Apache Hive, Hadoop, Dataflow, Dataproc
ETL, Ingestion, Workflow	SSIS, Talend, Informatica, Google Cloud Composer, Databricks, Snowflake
Data Modeling & Storage	PostgreSQL, MySQL, Oracle, MongoDB, SSMS
Containerization & CI/CD	Docker, Kubernetes, Helm, Azure DevOps, Jenkins
Monitoring & Security	Prometheus, Grafana, Stackdriver, KMS, IAM, VPC Service Controls, HIPAA Compliance
Data Governance	Great Expectations, Apache Atlas, GCP Data Catalog, Data Lineage, SCD Type 2, CDC Logic
Visualization & Reporting	Tableau, Looker, Power BI, Google Data Studio
Machine Learning Tools	Vertex AI, BigQuery ML, Scikit-learn, Pandas, NumPy, TensorFlow, PyTorch, OpenCV

TECHNICAL EXPERIENCE

Modak Analytics (Client: Humana Inc) - Data Engineer Hyderabad, India	July 2021 - July 2023
<ul style="list-style-type: none">Designed scalable event-driven data pipelines using Airflow, Pub/Sub, and GCS to ingest data from SSMS, Oracle, and MySQL - boosting throughput by 35% and reducing pipeline failures by 50%.Developed PySpark and Apache Beam ETL frameworks (Dataproc/Dataflow) to process 10+ TB/day of healthcare claims, cutting latency by 40% and enabling near real-time insights.Automated anomaly detection in ETL outputs using BigQuery ML and scheduled alerts with Cloud Functions and Slack API, improving issue resolution SLAs by 45%.Modeled analytical schemas (SCD Type 2, CDC) in BigQuery/PostgreSQL and built Looker Explores, accelerating \$5B+ claims analytics with sub-second query performance via materialized views.Integrated 5+ FHIR-compliant APIs into ingestion pipelines to support real-time processing of 1M+ clinical and eligibility records/day, enabling HIPAA-aligned data access and accelerating care analytics delivery by 30%.Containerized Spring Boot ingestion services and orchestrated deployments on GKE using Helm, autoscaling, and blue-green CI/CD (Cloud Build, Azure DevOps), slashing release time from 5 days to 8 hours.Implemented metadata-driven ETL orchestration and validation layers (Great Expectations, DLP APIs, Cloud Functions), reducing QA time by 60% and ensuring compliance with strict healthcare SLAs.Led security and governance initiatives including IAM hardening, KMS encryption, and data lineage via Apache Atlas-GCP Data Catalog integration; mentored 4 junior engineers on cloud-native best practices.Conducted cloud resource audits using GCP Recommender and Cost Table reports; implemented auto-scaling and idle resource shutdown policies, reducing monthly cloud spend by 22%.	

EDUCATION

George Mason University - M.S. in Computer Science Fairfax, VA (GPA: 3.87/4.0)	Aug 2023 - May 2025
SRM Institute of Science and Technology - B.Tech. in Electronics and Communication Engineering India (GPA: 3.3/4.0)	Aug 2017 - Apr 2021

ACADEMIC PROJECTS

Unmanned Aerial Vehicle-Decision Guidance System - UAV Trajectory Optimization Framework LINK	Jan 2025 - May 2025
<ul style="list-style-type: none">Engineered a modular UAV trajectory optimization framework using binary decision variables and 3D leg metric computation to support multi-objective path planning.Enabled KPI-driven routing (distance, energy, time) and visualized patrol patterns (circle, figure-8, rectangle); laid groundwork for PX4-ROS2-Gazebo simulation.	
PatriotPilot - NLP-Based University Chatbot LINK	Sep 2024 - Dec 2024
<ul style="list-style-type: none">Designed a domain-specific RAG pipeline using E5-Large-v2 embeddings and FAISS indexing for sub-100ms retrieval on academic Q&A.Achieved over 70% Top-1 accuracy on 500+ real student queries by fine-tuning Qwen 2.5 14B Instruct using LoRA and DeepSpeed ZeRO, optimizing model performance for academic Q&A tasks.	
GentopiaMason - Multi-Agent Research Assistant LINK	Sep 2024 - Dec 2024
<ul style="list-style-type: none">Extended Gentopia with LangGraph, agent memory, and tool-based decomposition to automate 85% of academic tasks, cutting manual effort by 60%.Deployed Mixtral/LLaMA 2 via vLLM on A100 MIG GPUs using Triton inference server with batching for high-throughput tasks.	
Survey Management System - Full Stack Web Application LINK	Sep 2024 - Dec 2024
<ul style="list-style-type: none">Developed a Vue.js-Spring Boot platform with Dockerized microservices, REST APIs, and Amazon RDS MySQL, deployed via Jenkins CI/CD on Kubernetes.Integrated Kafka for event logging and Terraform for IaC-driven cloud provisioning across 3 dev/test namespaces.	