YASWANTH K

+1(945) 286-1919 ◇ yaswanthchowdary200@gmail.com ◇ LinkedIn
Data Engineer ◇ Irving, TX ◇ Open to Relocation

SUMMARY

Data Engineer with 4+ years of experience designing, building, and optimizing large-scale ETL/ELT pipelines and real-time streaming solutions across AWS, GCP, and Snowflake. Proficient in Python, SQL, and PySpark with expertise in Apache Airflow, Kafka, and BigQuery. Proven success in cloud migrations, supply chain and inventory analytics, and performance optimization—including migration of 50+ workloads to AWS, cutting query times by 60% and infrastructure costs by 35%. Certified in AWS and GCP with domain expertise in retail and financial data platforms.

EDUCATION

Master of Science, Computer Science, University of North Texas — May 2025

SKILLS

Programming	Python, SQL, Java, Scala, C, Linux
Big Data	Apache Spark, PySpark, Hadoop, Kafka, Hive, Azure Synapse
ETL/Orchestration	Apache Airflow, DBT, Talend, Fivetran, Step Functions, Cloud Composer
Cloud Platforms	AWS (S3, Glue, Redshift, EC2, KMS), GCP (BigQuery, DataProc, Cloud Storage, Composer), Snowflake
Databases	Redshift, PostgreSQL, MySQL, Oracle, MongoDB, Cassandra, DynamoDB
DevOps	Git, GitHub Actions, GitLab, Docker, Kubernetes, Terraform, CloudFormation
Data Security	IAM, Encryption (AES-256, KMS), Data Masking, DLP, Access Controls, PCI-DSS
Modeling/Visualization	Dimensional Modeling, Star Schema, ERwin, Lucidchart, Power BI, QuickSight
Soft Skills	Agile/Scrum, Cross-functional Collaboration, Documentation, Stakeholder Communication

RELEVANT EXPERIENCE

Data Engineer — 01/2024 – Present
Nike — *Irving, TX*

- Designed and maintained **real-time and batch pipelines** on **AWS (S3, Glue, Redshift) and GCP BigQuery** to process sales, inventory, and supply chain data across global regions.
- Built scalable **dimensional models and star schemas** to enable unified reporting for merchandising, product, and customer analytics, reducing query times by 45%.
- Integrated multi-source data feeds (**POS systems, e-commerce transactions, supplier APIs**) using **PySpark and Airflow**, improving data freshness and reliability for downstream analytics.
- Automated **data quality checks and schema validation** using Great Expectations, cutting production issues by 30% and ensuring consistent KPIs across reporting teams.
- Partnered with digital analytics teams to deliver **Power BI and QuickSight dashboards** for product demand, sell-through rates, and customer engagement, accelerating business insights.
- Developed **time-series forecasting models** (Prophet) that improved accuracy of demand prediction by 87%, reducing overstock by 25% and supporting leaner inventory planning.
- Documented data models, lineage, and pipeline architecture in **Lucidchart and Confluence**, improving onboarding efficiency and cross-team collaboration by 50%.
- Implemented **data security controls** (IAM, KMS encryption, masking sensitive customer data) to safeguard PII and ensure compliance with internal security standards.

Data Engineer 02/2021 – 06/2023
DHL *Chennai, IN*

- Migrated **20+ high-volume SQL Server workloads to AWS Redshift using S3 staging layers**, reducing query latency by 40%.
- Deployed **Apache Airflow DAGs** to orchestrate batch and **real-time ETL pipelines** for logistics and finance systems, improving data accuracy by 30%.
- Integrated 7+ data sources **(APIs, flat files, third-party feeds)** using Python (Pandas, Boto3) to automate ETL pipelines.
- Conducted **data quality checks and schema validation** with Great Expectations, reducing production-level data issues by 35%.
- Automated infrastructure provisioning with **CloudFormation and implemented CI/CD via GitHub Actions**, reducing deployment time by 50%.
- Supported **PCI-DSS audits** by implementing encryption-at-rest and granular access logging, enabling full compliance with internal policies.

CERTIFICATIONS

Google Cloud Professional Data Engineer Aug 2022

AWS Certified Data Engineer – Specialty Apr 2023

PROJECTS

Real-Time Retail Demand Forecasting System. Created a real-time demand forecasting pipeline using PySpark, AWS Glue, Redshift, and Prophet to process over 10 million retail transactions. Generated weekly sales forecasts, reducing inventory overstock by 30% and improving demand planning accuracy. Integrated batch inference into Apache Airflow DAGs for automated scheduling and monitoring.

Global Supply Chain Data Lakehouse. Built a scalable retail data lakehouse on AWS using S3, Glue, and Athena to manage global supply chain and logistics data. Integrated Apache Kafka for real-time ingestion and applied schema versioning with optimized partitioning to reduce query latency by 40%. Delivered QuickSight dashboards to track order fulfillment and warehouse KPIs.

Customer Churn Prediction and Dashboard. Developed a customer churn prediction model using Python, Scikit-learn, and XGBoost, achieving 87% accuracy. Deployed the model using Flask within Docker containers and exposed REST APIs for real-time inference. Visualized churn metrics through Power BI dashboards segmented by customer behavior and time period to support targeted retention strategies.