HEMANTH AKKENAPALLY

+1 (667) 406-4046 ◇ Halethorpe, MD

akkenapallyhemanth@gmail.com ◇ linkedin.com/in/hemanthakkenapally ◇ GitHub: Hemanth-Akkenapally

SUMMARY

Data Engineer / Senior Data Analyst with 3+ years of experience designing and maintaining enterprise-scale ETL/ELT pipelines, data warehouses, and data lake architectures. Proficient in **Python, SQL, Apache Spark, and AWS services (S3, Glue, Lambda, EMR, Redshift)** for automated data ingestion, transformation, and reporting. Hands-on experience with **data modeling, DDL optimization, data partitioning, and security controls (encryption, access policies, audit logging)** in Snowflake, Redshift, and cloud data lake environments. Proven ability to collaborate with cross-functional stakeholders to publish and consume data at scale.

EDUCATION

MS in Data Science, University of Maryland, Baltimore County (UMBC)   GPA: 3.76   Dec 2024

SKILLS

Data Engineering & ETL	Python, SQL (DDL/DML/PL-SQL), Apache Spark, SparkSQL, HiveQL, dbt, ETL/ELT Pipelines, Apache Airflow, Kafka, Hadoop, Hive, SSIS, Azure Data Factory, Batch/Streaming Pipelines
AWS & Cloud	S3, Glue, Lambda, SNS, SQS, KMS, EMR, Redshift, SageMaker, EC2; Azure (Data Lake, Synapse); GCP BigQuery
Data Warehousing & Modeling	Snowflake, Redshift, BigQuery, Data Partitioning, Compression, Parallelization, Logical/Physical Data Modeling, Data Catalog & Metadata Management, Data Lake Architecture
Data Security & Governance	IAM, KMS Encryption, Column/Row-Level Security, OAuth, Audit Logging, Data Lineage, Access Controls, Validation Rules
Programming & Backend	Python (advanced), R, SAS, Bash, REST APIs, Flask, FastAPI, JSON, C/C++
ML & Analytics	scikit-learn, TensorFlow, PyTorch, XGBoost, NLP, Computer Vision, Time Series, Statistical Modeling
DevOps & Collaboration	Docker, Kubernetes, Git/GitHub, Jenkins, CI/CD, Terraform, Ansible, Agile/Scrum, JIRA, Confluence
Dashboards & BI	Tableau, Power BI, Streamlit, QuickSight, Looker, Plotly, Matplotlib, Excel

EXPERIENCE

Health Data Engineer & System Engineer   Jan 2025 – Present
SIGT LLC, Baltimore, MD

- Designed and deployed automated ETL pipelines in Python (Pandas, NumPy) and SQL to ingest, validate, and transform VNA, oscilloscope, and catheter telemetry data into a structured data lake, cutting manual data preparation by **30%** and increasing data availability by **40%**.
- Built REST APIs (Flask/FastAPI) for real-time RF sensor data ingestion and anomaly detection, with monitoring hooks for pipeline health, achieving **95%** system uptime and reducing unplanned data gaps.
- Implemented data partitioning, compression, and physical table optimization for time-series datasets to improve query performance and storage efficiency across analytical workloads.
- Developed and maintained data catalog and metadata documentation for lab datasets, enabling self-service data consumption by engineering and operations stakeholders.
- Containerized analytics services using Docker and Kubernetes, increasing data processing throughput by **65%** and enabling reliable, repeatable pipeline deployments.
- Created Power BI, Tableau, and Streamlit dashboards to monitor RF ablation parameters and safety thresholds, enabling faster data-driven decisions during lab sessions.

Data Engineer & Salesforce Developer   Sep 2021 – Jan 2023
Cognizant Technology Solutions, India

- Engineered ETL pipelines integrating Salesforce, SQL, and Snowflake to synchronize **100K+** monthly financial transactions and customer records, supporting downstream reporting and analytics for KeyBank.
- Designed and optimized Snowflake data warehouse objects using best-practice DDL, logical/physical data modeling, and data partitioning strategies to improve query performance and reduce compute costs.
- Implemented data security including field-level security (FLS), OAuth policies, row-level access rules, and audit-ready lineage views, lowering data integrity issues by **32%** and cutting audit preparation time by **20%**.
- Automated Salesforce workflows with SOQL, Apex, and Flow Builder, improving reporting efficiency by **55%** and reducing manual intervention in loan and payment processes.
- Built and maintained Tableau and Salesforce dashboards with drill-downs for portfolio performance, credit risk, and customer segmentation; optimized SQL aggregation logic to cut dashboard refresh times by **40%**.
- Documented data pipeline requirements, data models, and warehouse metadata to support internal business customers and cross-functional development teams.
- Applied DevOps practices (Git branching, sandbox testing, CI/CD) to Salesforce releases, achieving **95%** on-time deployment rate across a 10-member Agile team.

Data Analyst & Network Engineer Intern   May 2019 – May 2021
Cisco, Hyderabad, India

- Built ETL pipelines using SSIS and Azure Data Factory to automate extraction, transformation, and loading of network telemetry metrics into reporting databases, enhancing data freshness and monitoring coverage.
- Developed data models and SQL queries (DDL/DML) to structure and expose network performance datasets for analytical consumption, reducing reporting latency and manual effort by **40%**.
- Built Power BI and Tableau dashboards for operational KPIs, network performance, and capacity metrics, accelerating incident visibility and infrastructure planning.
- Performed quantitative analysis on capacity and latency data using Python and SQL to support network optimization and data-driven infrastructure decisions.

PROJECTS

TweetSense: Real-Time Sentiment Analysis Pipeline with BERT and AWS Python, PyTorch, AWS S3, Lambda, REST API   GitHub

- Built an end-to-end data pipeline ingesting **1.6M** tweets: raw data stored in AWS S3, processed via Lambda-triggered transformations, and scored by a fine-tuned BERT model for sentiment classification.
- Deployed the model via a Python REST API exposing live predictions from the X (Twitter) API, with logging and monitoring for class distributions, latency, and accuracy.
- Designed the pipeline for extensibility, with partitioned S3 storage layout and metadata documentation supporting future analytics consumption.

Automatic License Plate Recognition (ALPR) using YOLOv4 and Darknet Computer Vision, YOLOv4, OpenCV, Python   GitHub

- Developed an end-to-end ALPR pipeline: vehicle and plate detection using YOLOv4, followed by OCR-based character recognition for real-time video feeds.
- Collected and annotated license plate images, performed data augmentation, and fine-tuned YOLOv4 weights to improve detection accuracy on low-light and angled images.
- Designed evaluation scripts to compute precision/recall and frame-level throughput for smart surveillance and toll automation use cases.