POORNA SAI VEERA HARISH PEDAPUDI

Virginia (Open to Relocation) | +1 (571) 281-1394 | poornasaipedapudi@gmail.com | [LinkedIn](#) | [Website](#) | [GitHub](#)

Summary

Data Engineer with 5+ years of experience designing scalable data pipelines, deploying cloud-native ETL systems, and managing distributed data architectures across finance, SaaS, and education sectors. Experienced with AWS (S3, Glue, Lambda, Redshift), Azure Synapse, and PySpark to build ingestion frameworks and batch-processing workflows. Proficient in Python, SQL, and CI/CD practices, with strengths in data modeling, governance, and real-time pipeline orchestration. Pursuing opportunities to contribute as a Data Engineer in environments focused on cloud platforms, automation, and large-scale analytics infrastructure.

Technical Skills

Languages & Scripting: Python, SQL, R, Bash.

Cloud Platforms & Services: AWS (S3, Glue, Lambda, Redshift, Athena, EMR), Azure (ADF, Synapse, Data Lake), GCP (BigQuery).

Data Engineering Tools & Frameworks: Airflow, dbt, PySpark, SparkSQL, Hadoop, Kafka, Delta Lake.

ETL & Pipeline Orchestration: AWS Glue, Apache Airflow, Lambda, Azure Data Factory.

Databases & Storage: PostgreSQL, MySQL, MongoDB, Amazon RDS, Snowflake, DynamoDB, HDFS.

Data Modeling: Redshift, Snowflake, Star/Snowflake Schema, Data Lakes, Alteryx.

Visualization & Reporting: Power BI, Tableau, Looker, Excel.

DevOps, MLOps & ML Tooling: GitHub, Docker, MLflow, CI/CD, A/B Testing, Prompt Engineering, Scikit-learn, XGBoost.

Professional Experience

Data Engineer, Indukuri Information Services Inc (Volunteer), Remote　　　　　　　　　　**May 2024 − Present**

- Spearheaded automation of Airflow DAGs and data ingestion from S3 to Redshift, reducing manual effort by 75% while coordinating cross-team data handoffs for 2M+ daily logs
- Engineered PySpark jobs transforming 5TB+ weekly raw data into analytical layers, improving model input and boosting classification AUC by approximately 12%
- Implemented secure Redshift schemas with CI/CD via GitHub Actions, reducing release delays from 5 days to 8 hours and increasing deployment traceability
- Consolidated 15 fragmented dashboards into 3 Power BI visualizations, improving leadership visibility and reducing metric redundancy by more than 70%
- Created flow documentation and LucidCharts for 10+ pipelines, reducing onboarding time by 40% and improving cross-team alignment across key stakeholders

Data Analyst, Hildreth Institute, Fairfax, VA　　　　　　　　　　**Jan 2024 – May 2024**

- Delivered 6 interactive Tableau dashboards using Redshift KPIs to analyze 250K+ student records, enhancing decision-making across 3K+ higher-ed institutions
- Eliminated manual steps by deploying Python scripts, and S3-based pipelines, saving 25+ analyst hours per reporting cycle
- Directed integration of dbt models and Lambda automation for quarterly S3-based reporting, doubling delivery speed and ensuring audit compliance across teams

Data Engineer, Tritech Software Pvt Ltd, Hyderabad, India　　　　　　　　　　**Jan 2020 – Dec 2022**

- Architected ETL pipelines in Python and AWS Glue to ingest 1M+ ERP records daily, reducing manual reporting delays from 12 hours to under 2 hours
- Configured PySpark transformations on EMR to clean 800GB/month of transactional data, improving pipeline efficiency by 30% and reducing memory-related failures
- Designed S3-based data lakes with Delta format and Hive partitioning, boosting Athena query performance by 45% and reducing storage costs by $9K monthly
- Established granular IAM access policies for Redshift layers and S3 buckets, improving security compliance and reducing audit prep time by nearly 50%
- Led development of 30+ modular dbt models and reusable macros to standardize reporting logic across 5 departments, reducing BI dependencies by 40%

Data Analyst, Tritech Software Pvt Ltd (Hyderabad, India)　　　　　　　　　　**May 2019 – Jan 2020**

- Queried 500K+ historical sales records to uncover seasonal demand patterns, enabling supply teams to reduce overstocking by 20% across 3 product categories
- Built Tableau dashboards with 15 KPIs across 5 regions, improving business insight and reducing reliance on outdated Excel-based reporting systems
- Facilitated documentation of schema logic and mappings across systems, ensuring a smooth Redshift migration and saving 60+ QA hours

Projects

- See real-world projects on data pipelines and analytics in [portfolio projects](#) section

Education

Masters, Data Analytics Engineering, George Mason University (Fairfax, VA)	**Jan 2023 – May 2024**
Bachelors, Electronics and Communication Engineering, Malaviya National Institute of Technology (Jaipur, India)	**Jul 2016 – May 2020**