



Lakkaraju Anjaneya Vara Prasad

Programmer Analyst

✉ anjaneyavaraprasad01@gmail.com ☎ 9573294994 📍 hyderabad,India

🌐 <https://www.linkedin.com/in/lakkaraju-anjaneya-vara-prasad-00070b235> 📅 04/08/2001 🇮🇳 Indian

👤 **Headline**

Data Engineer | AWS Glue – EMR – Redshift | ETL & Data Modeling | SQL – Python – Spark | Azure & GCP Familiar

👤 **Profile**

Motivated Data Engineer with 1.2 years of experience supporting cloud-based data pipelines and analytics solutions. Contributed to a banking analytics project for Fifth Third Bank, working with AWS Glue, EMR (PySpark), and Redshift to process and transform transactional data. Skilled in SQL, Python, and Spark for data modeling, transformation, and performance tuning. Delivered datasets that improved reporting accuracy by ~10% and reduced manual effort by ~15%. Familiar with Azure and GCP platforms, with working knowledge of services like Azure Data Factory, Synapse, BigQuery, and Dataflow. Quick learner with a focus on building reliable, secure, and cost-efficient data workflows.

🎓 **Education**

Bachelor of Technology (B.Tech), Vasireddy Venkatadri Institute of Technology 06/2019 – 05/2023
CGPA : 8.0 Guntur, India

👤 **Professional Experience**

Programmer Analyst, Cognizant Technology Solutions 12/2023 – Present
hyderabad, India

- Assisted in building and maintaining ETL pipelines processing 1M+ daily banking transactions using **AWS Glue** and **EMR (PySpark)**, integrating data from PostgreSQL, MySQL, and flat files
- Developed SQL and Python scripts for data cleansing and transformation, improving reporting accuracy by ~10%
- Supported data modeling in **Amazon Redshift** for fraud analytics and customer segmentation dashboards
- Helped automate pipeline orchestration using **AWS Lambda** and **Step Functions**, reducing manual intervention by ~15%
- Contributed to Spark job tuning using partitioning and caching, resulting in ~20% faster execution times
- Delivered curated datasets for **Power BI dashboards** used by risk and compliance teams
- Monitored EMR clusters and collaborated on cost optimization strategies, reducing compute overhead by ~10%
- Worked with analysts, QA, and cloud architects to ensure secure, compliant, and high-quality data delivery

- Explored Azure and GCP services to broaden cloud platform understanding and support cross-platform comparisons

Skills

Technical Skills:

- **Cloud Platforms:** AWS (Glue, EMR, Redshift, Lambda, S3), Azure (Data Factory, Synapse), GCP (BigQuery, Dataflow)
- **ETL & Data Integration:** AWS Glue, Lambda, Step Functions
- **Big Data & Distributed Processing:** EMR (Spark), Hadoop, Hive, Sqoop, HDFS
- **Data Modeling & Warehousing:** Amazon Redshift, PostgreSQL, MySQL
- **Programming & Query Languages:** SQL, Python, SparkSQL, HiveQL
- **Reporting & Visualization:** Power BI, curated datasets, stakeholder dashboards
- **Security & Governance:** IAM, KMS, RBAC, Key Vault
- **Monitoring & Optimization:** CloudWatch, EMR tuning, job performance tracking

Soft Skills:

- Communication
- Teamwork
- Problem-solving
- Attention to detail
- Time management

Languages

- English
- Telugu
- Hindi

Certificates

Microsoft Certified: Azure Data

Engineer Associate (DP-203)

(AWS certification in progress – targeting AWS Certified Data Analytics or Solutions Architect)

Mini Projects

Streaming state management for user session tracking using pyspark:

This project leverages PySpark Structured Streaming to implement real-time session tracking and stateful analytics. It processes a continuous stream of employee activity events—such as logins, clicks, and logouts—to dynamically manage session states. By applying techniques like sessionization, stateful aggregation, and windowed operations, the system provides live insights into session durations, user engagement, and cumulative actions. The project showcases the power of streaming state management in operational analytics, enabling responsive dashboards and real-time decision-making across domains like HR, security, and product usage.

Cloud Data Orchestration: From Buckets To BigQuery With Dataflow and Cloud functions

This project automates the end-to-end pipeline for ingesting, processing, and loading data from Google Cloud Storage into BigQuery using Cloud Functions and Dataflow. The workflow is event-driven: when a new file is uploaded to a bucket, a Cloud Function is triggered to launch a Dataflow job that transforms and loads the data into BigQuery. The solution demonstrates scalable and efficient cloud-native orchestration, highlighting seamless integration across Google Cloud Platform (GCP) services. It emphasizes real-time responsiveness, modular architecture, and cost-effective data movement.