

# Lokesh Reddy Lingala

+1 4846454624 • San Jose, CA • [Portfolio](#) • [lokeshreddy.de@gmail.com](mailto:lokeshreddy.de@gmail.com) • [LinkedIn](#) • [GitHub](#)

## SUMMARY

**Data Engineer** with 2+ years of experience designing **scalable, idempotent** ETL pipelines, data modeling, and optimizing big data processing while ensuring **data quality, integrity, and lineage** across workflows. Proficient in Python, PySpark, SQL, and leveraging cloud(AWS, Snowflake). Focused on **not only delivering data** but ensuring it drives meaningful, data-driven decisions, empowering teams to take informed actions and drive business outcomes.

## WORK EXPERIENCE

### Research Assistant

Buffalo, NY

### University At Buffalo

Feb 2024 - Dec 2024

- Modeled a scalable **citation network** with **Neo4j**, modeling 1M+ research papers and co-authorship patterns to identify influential researchers. Batch **weekly ETL** to update database with new publications.
- Optimized **graph traversal queries**, cutting query time by 50%, and empowered university research teams with data-driven insights for academic impact assessment and identifying key research trends.

### Data Engineer

Hyd, India

### SABIC - Saudi Basic Industries Corporation

Jun 2021 - Jul 2023

- Designed and implemented scalable data pipelines using **AWS(S3, Glue, Lambda and Redshift)**, automating **ETL** with **PySpark** and **SQL** to process large datasets, provisioning infrastructure with **Terraform**.
- Set up and fine-tuned Apache **Spark** clusters on **AWS EMR** to process 10+ TBs of customer and sales data, downstreaming it to the Data Science team to develop risk forecasting and revenue prediction models.
- Contributed to **migrating on-premises** databases to **AWS** by setting up **RDS** instances, ensuring smooth integration with **DMS** and **SCT**, Supporting team to complete migration while minimize downtime.
- Redefined a **CI/CD** process for a critical data pipeline using **GitHub Actions** and **Airflow**, automating schema validation and dependency checks, reducing deployment time from 1 hours to 15 minutes.
- Built and maintained **50+ fact and dimension tables** with **dbt(Data Build Tool)**, achieving a **30%** reduction in storage costs via cumulative tables and ensuring historical accuracy using **SCD(Type 2)** techniques.
- Applied advanced **SQL** skills to support teams with **ad-hoc** queries, accelerating insights generation and pre-aggregating Key Performance Indicators(**KPI**) to streamline BI reporting and enhance decision-making.
- Implemented a robust **data quality framework** using **Glue** and **Airflow** with 12 validation rules, enabling finance team improve reporting accuracy, prevent compliance issues, and reduce manual validation efforts by 40%.
- Acquired knowledge in Event-driven arch, ELT, MIDAS, Jenkins, Iceberg, Docker, Kubernetes and REST API's.

## TECHNICAL SKILLS

**Programming:** Python, Pyspark, SQL, HiveQL, Java, HTML, CSS, R, and C Programming.

**Data Warehousing:** Redshift, Snowflake, ClickHouse, BigQuery, PostgreSQL, MySQL, Neo4j, DynamoDB, MongoDB.

**AWS Services:** S3, Glue, Lambda, EC2, Athena, RDS, EMR, Kinesis, and DMS.

**Libraries and Tools:** Pandas, Power BI, Kafka, Spark, Airflow, Flink, Iceberg, Git, dbt, Docker and VS Code.

**Soft Skills:** Teamwork, Time management, Problem-solving, strong verbal, and written communication skills.

## PROJECTS

### End-to-End Real-Time Data Pipeline for Autonomous Vehicles [Link]

Jan 2025

- Developed a real-time pipeline with **Kafka**, **Flink**, and **TimescaleDB**, achieving **sub-100ms** latency for vehicle telemetry. Containerized with **Docker** and created **Streamlit dashboards** for telemetry and alerts.

### Real-Time Health Data Pipeline [Link]

May 2024

- Built a real-time pipeline with **AWS Kinesis**, **Lambda**, and **Snowflake** to process smartwatch records daily with **sub-5s latency**, improving efficiency by **35%**. Integrated **Snowpipe** and **Power BI** for real-time analytics.

## EDUCATION

### University At Buffalo - SUNY

Buffalo, NY, USA

Master's in Data Science CGPA: 3.8

Aug 2023 - Dec 2024

## CERTIFICATIONS

- MongoDB Certified Associate Data Modeler — **MongoDB, Inc.** — Jan 2025
- AWS Certified Data Engineer – Associate — **Amazon Web Services (AWS)** — Aug 2024