**Nikhitha Karimsetty**

+1 (512) 2149718 | [nikhitha.karimsetty@gmail.com](mailto:nikhitha.karimsetty@gmail.com)

LinkedIn: [https://www.linkedin.com/in/nikhithakb5a991175/](https://www.linkedin.com/in/nikhitha-k-b5a991175/)

GitHub: <https://github.com/nkarimsetty>   
**Professional Summary**

Detail oriented and performance driven Data Engineer with 4+ years of experience building data pipelines, lakes, and analytical frameworks. Proficient in Python, SQL, Java, Scala, Apache Spark, Kafka, Hive, Airflow, NiFi, DBT, Snowflake, Redshift, Azure Data Factory, and Databricks. Designed scalable ETL workflows and automated data validation using Jenkins, Docker, Terraform, and CI/CD pipelines. Expert in data modeling (Star/Snowflake Schema), cloud services (AWS/GCP/Azure), REST APIs, and data governance principles. Strong communication, collaboration, and problem-solving skills with a focus on delivering business impact through secure, reliable, and high-quality data solutions.

**Education**

Master of Science in Computer Science – University of Central Oklahoma | Aug 2023 – May 2025

Coursework: Advanced Programming in C++, Algorithms, Computing Practicum, Front-End Web, Software Engineering I & II, Cloud Web Apps, Database Systems, Mobile Apps, Theory of Computing, Enterprise Web, Cyber Infrastructure & Cloud.

**Skills & Technologies**

**Languages & Scripting:** Python, SQL, Java, Scala, Bash  
**Big Data & Orchestration:** Apache Spark, Apache Hadoop, Apache Airflow, Apache NiFi, Kafka, DBT, Google Cloud Composer  
**Cloud Platforms & Tools:** AWS (S3, Glue, Redshift, Lambda, EMR), Azure (Data Factory, Synapse, Data Lake), GCP (BigQuery, Dataflow, Pub/Sub)  
**Data Storage & Modeling:** PostgreSQL, MySQL, Oracle, SQL Server, MongoDB, Cassandra, DynamoDB, Snowflake  
**DevOps & Automation:** Jenkins, Git, Docker, Kubernetes, Terraform, CloudFormation, CI/CD Pipelines  
**Data Management:** Data Warehousing, Data Lakes, Star/Snowflake Schema, Data Lineage, Data Security, Data Quality, Governance  
**Visualization:** Power BI, Tableau, Looker  
**Others:** REST APIs, ELK Stack, Agile Methodologie

**Professional Experience**

**Sr. Data Engineer**

**Infosys Nov 2020 – Jul 2023**

* Engineered Kafka based data streaming pipelines integrated with Hive, NiFi, and DBT to support real time analytics for key clients
* Built scalable ingestion frameworks on AWS EMR, Lambda, and Redshift serving 20+ production data applications
* Developed REST APIs and custom data services to transform raw operational data for enterprise analytics
* Created modular Terraform IaC for provisioning Snowflake DWH and associated monitoring pipelines
* Migrated legacy batch workflows to Apache Airflow and Azure Data Factory, reducing SLA breaches by 70%
* Integrated Jenkins and Git for automated unit testing, deployment validation, and logging for data quality
* Enhanced data security using IAM roles and encryption at rest practices across AWS and Azure cloud platforms
* Performed regression testing and validation on over 2TB of raw/structured data for accurate KPI dashboards
* Coached junior engineers on CI/CD, Git workflows, and Apache Spark best practices in weekly team syncs

**Data Engineer Intern**

**Onspec Technology Solutions Apr 2019 – oct 2020**

* Built and maintained scalable data pipelines using Spark, Python, and SQL on AWSEMR and AzureDataFactory for large-scale dataprocessing.
* Automated ETL workflows with Airflow, Glue, and DBT, processing data from APIs, Kafka, and cloud sources for real**-**timeanalytics.
* Modeled data warehouses using Star and Snowflakeschemas in Snowflake, Redshift, and BigQuery with a focus on datawarehousing.
* Ingested and transformed data across S3, AzureDataLake, and GCP, enabling unified reporting and dataintegration.
* Performed SQL optimization and data extraction on PostgreSQL, MySQL, and MongoDB for ad hoc dataanalysis.
* Built dashboards in PowerBI, Tableau, and Looker for business reporting and datavisualization.
* Applied dataquality, security, and governance checks to ensure trusted analytics pipelines and compliance.
* Used Git, Docker, and Jenkins for CI/CD in a collaborative Agile environment with DevOps practices.
* Documented data lineage and collaborated with cross-functional teams in weekly sprint reviews using JIRA.

**Certifications**

Google Cloud Certified: Professional Data Engineer  
 AWS Certified Data Analytics – Specialty  
 Microsoft Certified: Azure Data Engineer Associate  
 Databricks Lakehouse Fundamentals  
 Confluent Certified Developer for Apache Kafka

**Projects**

Serverless Participation Verification System – AWS Lambda, Rekognition, Textract, DynamoDB, React

Data Lake Ingestion Framework – Kafka, Apache NiFi, Hive, Redshift