

# VAMSI THOKALA

## Data Engineer

913 368 7174 | [✉️ thevamsithokala@outlook.com](mailto:thevamsithokala@outlook.com) | [🌐 https://www.linkedin.com/in/vamsi-thokala-8255052a2/](https://www.linkedin.com/in/vamsi-thokala-8255052a2/)

### PROFESSIONAL SUMMARY

Senior Data Engineer with 8+ years of experience modernizing enterprise data platforms and leading large-scale ETL migrations from legacy tooling into Cloud Lakehouse. Hands on experience in PySpark, Spark SQL, Delta Lake, and workflow orchestration, experience applying Infrastructure-as-Code practices to build and manage cloud data infrastructure, modernizing complex pipelines into scalable, production-grade solutions. Known for operating independently, maintaining playbook during migration, and documenting systems to reduce operational risk. Experienced across Azure and AWS environments, supporting analytics, reporting, and regulated workloads with high reliability and minimal disruption.

### PROFESSIONAL EXPERIENCE

#### Rewards Network

##### *Senior Data Engineer*

Dec 2025 – Present

- Owned a major BI migration by moving Domo sources to Redshift and integrating Salesforce through reusable Python and PySpark components.
- Migrated Domo dataset sources from SQL Server to Amazon Redshift, validating row-level accuracy and metric consistency across old and new systems.
- Built reusable ingestion and transformation components in Python and PySpark, enabling faster onboarding of new data sources.
- Partnered closely with analytics and business stakeholders to align migrated datasets with existing dashboards and KPIs.
- Authored operational runbooks and validation checklists to support production stability and reduce single-engineer dependency.

#### Alaska Airlines

##### *Senior Data Engineer*

Sep 2025- Dec 2025

- Supported enterprise data modernization initiatives by improving pipeline reliability, dataset quality, and operational visibility for analytics workloads.
- Built Databricks notebooks and production Jobs to migrate SQL-centric logic into scalable PySpark workflows with standardized logging and error handling.
- Contributed to enterprise-scale data modernization initiatives focused on migrating SQL-heavy analytics workloads into Azure Databricks.
- Analyzed legacy SQL-centric ETL workflows and translated them into Databricks PySpark pipelines optimized for scale and reliability.
- Built production-grade Databricks notebooks and Jobs with standardized logging, error handling, and data quality checks.
- Implemented Delta Lake-based transformations to ensure data consistency and reproducibility across environments.
- Worked directly with analytics teams to validate migrated datasets against legacy outputs, ensuring full metric parity.
- Improved operational visibility by introducing structured pipeline documentation and troubleshooting guides.

#### First Fidelity Bank – Oklahoma

##### *Senior Data Engineer*

April 2024 – July 2025

- Built and scaled hybrid data pipelines integrating MS SQL Server with AWS services (PostgreSQL, Athena, S3, EC2, VPC) to support regulatory reporting and risk analytics under FINRA and cybersecurity compliance requirements.
- Developed reusable ETL/ELT workflows using Python, PySpark, and Apache Airflow to transform financial transaction data into curated canonical models, reducing data errors by ~35%.
- Expanded self-service orchestration frameworks enabling analysts and engineers to build and run pipelines with minimal platform dependency, aligned with data governance standards.
- Designed and maintained metadata and cataloging layers to improve data discovery, enforce access controls, and standardize transformations across systems.
- Built containerized pipelines using Docker and Kubernetes, applying CI/CD and DevOps practices to streamline deployment and improve reliability.

- Partnered with ML, data science, and finance teams to curate canonical datasets and onboard analytical workloads into the lakehouse environment.
- Researched and introduced modern tools (Databricks Lakehouse, dbt, Kafka) to improve scalability, governance, and architectural consistency.
- Automated monitoring, alerting, and production support using GitLab CI/CD, Ansible, and Azure Monitor to improve observability and issue resolution.
- Authored detailed technical documentation covering ETL frameworks, architectural decisions, and complex project implementations.

## **Chesapeake Energy Corporation– Oklahoma, USA**

### **Data Engineer**

*Sep 2023 – March 2024*

- Implemented centralized configuration and repository management practices to streamline pipeline development and deployment.
- Developed real-time streaming ingestion using Pub/Sub and Dataflow alongside batch ETL with Dataproc, achieving sub-minute latency for SCADA monitoring.
- Optimized BigQuery workloads using partitioning, clustering, and materialized views, reducing reporting latency by ~45%.
- Designed architecture standards and technical documentation supporting data curation and governance initiatives.
- Participated in architecture design reviews and sprint planning, aligning technical priorities with business requirements.

## **University of Oklahoma – Oklahoma, USA**

### **Graduate Assistantship**

*Aug 2022 – Sep 2023*

- Built event-driven pipelines using Pub/Sub and Dataflow to stream patient admission and discharge events into near real-time dashboards.
- Developed analytical datasets in BigQuery and dbt, curating 15M+ healthcare records into reusable research-ready datasets.
- Automated ingestion workflows using Airflow and Terraform, significantly reducing manual intervention.
- Worked in agile sprints, contributing to design, prioritization, and troubleshooting of complex data workflows.

## **System Info Solutions**

### **Big Data Engineer**

*Aug 2020 – Mar 2022*

- Developed Spark-based data ingestion and transformation jobs on Hadoop/HDFS, consolidating application and transaction data into centralized data platforms for reporting and analytics.
- Tuned distributed jobs and SQL queries to improve performance and stability across large datasets, supporting downstream dashboards and reports.

## **Indian Eagle Pvt. Ltd.**

### **Data Associate**

*May 2016 – Jul 2020*

- Built and supported ETL workflows using Spark, Python, and SQL to integrate web, booking, and transaction data into analytical data stores for marketing and operations teams.
- Collaborated with analysts to define data models and metrics, improving reporting accuracy for customer and revenue analytics.

## **TECHNICAL SKILLS**

- AWS (S3, Glue, Redshift, EMR, Lambda, Step Functions, Athena, EC2, VPC, CloudFormation), Azure (Databricks, Synapse, Data Factory, Key Vault), GCP (BigQuery, Dataflow, Pub/Sub), Snowflake, Hadoop, Hive, Presto, Kafka
- Databases & Data Warehousing:
- PostgreSQL, Oracle, SQL Server, MySQL, Cassandra, DynamoDB, Data Warehousing, Data Modeling (Star Schema, 3NF, Data Vault)
- ETL / ELT & Data Integration:
- Apache Spark (PySpark, Scala), Airflow, dbt, Glue Workflows, SSIS, Informatica, Talend, CDC, Data Transformation
- Programming & Scripting:

- Python, SQL, Scala, Java, Shell Scripting, Node.js
- BI & Analytics:
  - Tableau, Power BI, Looker, Domo, Qlik, Excel (Pivot Tables, Power Query)

## EDUCATION

- **Master's degree in data science & Analytics**, University of Oklahoma
- **Bachelor's degree in computer science**, Indian Institute of Technology (ISM) Dhanbad