SATHVIKA KOLISETTY

+1 (937) 400-9988 | sathvika.kolisetty09@gmail.com | Dayton, OH, USA | LinkedIn | Portfolio

PROFESSIONAL SUMMARY

Results-driven Data Engineer with 4+ years of experience designing, building, and optimizing data pipelines and analytics platforms across **AWS and GCP environments**, with domain expertise in **finance, insurance, and healthcare**. Proven success in reducing processing latency by 60%, automating 90%+ of workflows, and improving query performance by 45% through advanced ELT design, scalable data modeling, and infrastructure automation. Proficient in **Python, SQL, Spark, Databricks, dbt, Airflow, and Snowflake**, with deep hands-on experience in real-time streaming (Kafka, Kinesis, Flink) and CI/CD using Terraform. Known for delivering secure, production-grade solutions in agile teams that drive both operational efficiency and strategic impact. Certified **Databricks Data Engineer Professional**, **AWS Solutions Architect Professional**, and **AWS Data Engineer Associate**.

SKILLS

Programming & Scripting: Python, SQL, Scala, Java, Bash, C/C++
Cloud Platforms: AWS (Glue, Lambda, Redshift, Kinesis, S3, Step Functions, DMS, API Gateway, RDS, IAM, CloudWatch), GCP (BigQuery, Dataflow, Pub/Sub, Composer), Azure (Synapse, Data Factory)
Data Engineering & Orchestration: Databricks, Spark (PySpark), Airflow, dbt, Flink, Kafka, Kinesis
Modeling & Data Warehousing: Star Schema, Snowflake Schema, Fact-Dimension Modeling; Snowflake, Redshift, BigQuery, Synapse, PostgreSQL, MySQL, SQL Server, DynamoDB, MongoDB
Big Data & Real-time Streaming: Delta Lake, Apache Hive, Trino, Cassandra, Kafka, Flink, Spark Structured Streaming, Hadoop
APIs & Integration: REST APIs, FastAPI, AWS API Gateway, Lambda Functions, Glue Jobs
Infrastructure & DevOps: Terraform, CloudFormation, Jenkins, GitLab CI/CD, Docker, Kubernetes (EKS, GKE)
Monitoring & Security: CloudWatch, Datadog, Prometheus, IAM, Lake Formation, KMS, Great Expectations
Data Visualization & Analytics: Tableau, Power BI, Looker, Amazon QuickSight

PROFESSIONAL EXPERIENCE

Swift   **Remote**
Cloud Data Engineer   *May 2023 - Present*
- Led the end-to-end architecture of real-time data streaming pipelines using Kinesis, Flink, Lambda, and Redshift, reducing latency by 60% and enabling sub-minute operational insights.
- Migrated 10+ pipelines to AWS-native ELT stacks (Glue, Redshift, Snowflake) with zero downtime, automating 90%+ of workflows using Terraform and CI/CD pipelines.
- Built and optimized Star/Snowflake schemas and real-time data models to support self-service analytics, improving query speed by 45%.
- Built and deployed RESTful APIs using FastAPI, Lambda, and API Gateway, streamlining data delivery to analytics and ML teams, accelerating model development and dashboard refresh cycles by 40%.
- Spearheaded cost optimization across compute and storage (Redshift WLM, Snowflake scaling, S3 tiering), cutting infrastructure costs by 30%.
- Collaborated closely with product, analytics, and data science teams to deliver business-aligned solutions in a fast-paced, cross-functional environment.
- Ensured data governance and compliance using IAM, KMS, and Lake Formation, ensuring SOC2-aligned security for sensitive datasets.
- Took initiative to mentor new team members on CI/CD best practices, contributing to faster onboarding and higher team velocity.

WTW   **Remote**
Data Engineer   *April 2020 - July 2022*
- Modernized legacy pipelines by migrating batch jobs to Apache Spark (PySpark) on Databricks, reducing processing time by 60% and improving scalability for large datasets.
- Designed and maintained complex ELT workflows using Airflow, dbt, Glue, and Python, improving data reliability and integration speed by 50%.
- Built streaming data pipelines using Kafka, Kinesis, and Flink to power real-time insights for critical reporting.
- Developed automated data validation and alerting frameworks using SQL, Python, and dbt, improving data trust across BI platforms.
- Led a company-wide data quality initiative, standardizing metrics and building dashboards that drove 3x adoption of analytics tools by business users.

- Optimized Spark cluster configurations and Databricks resource utilization, cutting job costs by 25% while maintaining SLAs for large-scale batch workflows.
- Partnered with BI and product teams to design intuitive data models and dashboards, enabling non-technical users to generate actionable insights.
- Ensured data security, governance, and compliance through IAM access controls, encryption, and audit logging, aligning with SOC2 and GDPR standards.

PROJECTS

Scalable Clickstream Data Pipeline for User Behavior Analytics
Tech Stack: Apache Kafka, Apache Spark (Structured Streaming), Delta Lake, Apache Hive, Apache Hudi, Trino, Airflow, Databricks, AWS S3, Redshift, dbt, Terraform
- Engineered a Spark-based real-time pipeline ingesting 5M+ daily events from Kafka, transforming and storing data in Delta Lake and Hudi for unified batch and stream access.
- Enabled low-latency ad-hoc analysis with Trino and Hive Metastore integration, reducing data discovery time by 50% for analytics teams.
- Automated data quality checks and schema validations using dbt and Airflow, improving pipeline reliability and business confidence.

Healthcare Data Lakehouse with Real-Time Analytics
Tech Stack: Azure Data Factory, Delta Lake, Synapse, Power BI, Azure ML, Terraform
- Built a HIPAA-compliant Azure lakehouse pipeline to ingest, model, and visualize healthcare claims data in real-time.
- Deployed predictive models in Azure ML to identify high-risk patient readmissions with 87% accuracy.
- Accelerated analytics with 40% faster dashboard performance and 35% reduced data load times.

Cross-Cloud GenAI Recommendation Engine
Tech Stack: Snowflake, AWS Kinesis, AWS Glue, Lambda, GCP Vertex AI, S3, FastAPI, Streamlit
- Developed a real-time GenAI-powered recommendation engine leveraging behavior data ingested via Kinesis and modeled with Vertex AI.
- Centralized data in Snowflake with efficient ELT via AWS Glue and deployed a Streamlit-based frontend for live demo.
- Boosted recommendation accuracy by 30% and cut Snowflake compute costs by 25% using partitioned tables and materialized views.

CERTIFICATIONS

- **Databricks Certified Data Engineer Professional** – *Issued 2025*
- **AWS Certified Solutions Architect – Professional** – *Issued 2025*
- **AWS Certified Data Engineer – Associate** – *Issued 2025*

EDUCATION

SRM Institute of Science and Technology
Bachelor's, Computer Science   *GPA: 3.7*

University of Dayton
Master's, Computer Science   *GPA: 3.9*