

Sai Venkata Sahith Reddy Kadapaiahgari

Frankfurt, Germany | +49 17634609021 | @mail | @linkedin | @github

PROFESSIONAL SUMMARY

Data Engineer with 3+ years of experience designing and scaling cloud-native data platforms across Azure and AWS. Specialized in building **ELT** pipelines, data warehouses, and analytics-ready datasets using **Python, SQL, Spark, dbt**, and **Databricks**. Proven track record of processing TB-scale data, optimizing performance, and enabling BI and machine learning use cases in enterprise environments.

TECHNICAL SKILLS

Core Languages : Python, SQL, PySpark

Big Data & Processing : Apache Spark, Databricks

Orchestration & Streaming : Apache Airflow, Apache Kafka

Databases & Warehouses : Snowflake, Amazon Redshift, PostgreSQL, MySQL, Oracle, MongoDB

Cloud Platforms : Microsoft Azure (Data Factory, Synapse, ADLS), AWS (S3, Glue, Lambda, Redshift), Microsoft Fabric, Google Cloud Platform (BigQuery)

BI Tools : Tableau, Power BI, Looker

DevOps & Collaboration : Git, GitHub, Azure DevOps, CI/CD, Agile/Scrum

EXPERIENCE

Working Student Data Engineer | Qualifyze

September 2023 – June 2025

- Built and optimized ELT pipelines using **dbt** to transform data directly in **Amazon Redshift**, integrating multi-source data and accelerating analytics delivery by 15%.
- Enhanced data quality and performance by redesigning schemas and optimizing **SQL**, resulting in 35% faster queries and more reliable insights.
- Curated and delivered production-ready feature datasets for machine learning models (TensorFlow, Scikit-learn), reducing data preparation time for data scientists by an estimated 10 hours per week.
- Developed interactive dashboards in **Looker** and **Tableau** that cut reporting time by 30% and improved executive decision-making speed by 20%.
- Enhanced team collaboration through **GitHub** version control and Agile methodologies, improving project execution by 10%.

Data Engineer | Tata Consultancy Services

September 2020 – July 2022

- As a TCS consultant embedded with Ericsson, engineered and maintained their Azure data platform handling global telecommunications data.
- Engineered enterprise-scale data pipelines processing 1TB+ of daily telemetry data using **Azure Data Factory** and **Databricks**, reducing end-to-end latency by 20%.
- Optimized SQL queries and stored procedures, improving performance by 35% and reducing infrastructure costs.
- Automated deployment workflows with **Azure DevOps CI/CD**, eliminating 40+ hours of manual effort per month and accelerating release cycles.
- Designed scalable data warehouse models (**star/snowflake schemas**), which boosted BI dashboard performance and reduced refresh times.
- Established data quality frameworks and monitoring alerts for critical Ericsson data pipelines, reducing data incidents by 25% and ensuring reliable data for business-critical decisions.
- Partnered with cross-functional teams to deliver regulatory and financial reporting solutions, directly supporting 200+ business users across departments.

CREDENTIALS

Microsoft Certified Fabric Data Engineer Associate | Microsoft [Credential]

June 2025

Data Engineer Associate | DataCamp [Credential]

June 2025

Creating Dashboards and Storytelling with Tableau | Coursera [Credential]

May 2023

EDUCATION

University of Koblenz
RGM College of Engineering

MS in Web and Data Science | Germany
BS in Computer Science and Engineering | India

October 2022 – March 2026
June 2016 – March 2020

PROJECTS

Azure Odyssey: End-to-End Data Flow | ETL, Data Factory, Databricks, Synapse, PySpark. [GitHub]

June 2025

- Designed and implemented a production-grade ELT data platform using Azure Data Factory, Databricks, and Synapse.
- Implemented medallion architecture (Bronze/Silver/Gold) using PySpark to support scalable analytics.
- Reduced pipeline runtime by 25% through parameterization and optimized Spark jobs, enabling faster self-service BI in Power BI.