

SAMBA KUMAR GOLI

📞 +91-8801602981 📩 kumarsamba07@gmail.com 💬 linkedin.com/in/sambakumargoli

Summary

Data Engineer with 3 years of hands-on experience designing and building scalable data pipelines on Azure using Databricks, ADF, ADLS Gen2, and Azure SQL. Strong practical background in PySpark, Delta Lake, and data modelling. Experienced in developing and optimizing ETL/ELT workflows, implementing CI/CD with Azure DevOps, and supporting production pipelines with solid monitoring and reliability practices. Skilled at improving pipeline performance, strengthening data quality, and delivering analytics-ready datasets across Medallion architecture (Bronze-Silver-Gold).

Skills

Programming & Querying: SQL, Python (Pandas, data validation, automation scripts), PySpark

Data Engineering & Processing: ETL/ELT Pipeline Design, Data Cleaning & Validation, Data Quality Checks, Descriptive Statistics for Profiling, Batch & Stream Processing (Spark), Workflow Orchestration (ADF).

Data Modeling & Storage: Medallion Architecture, Data Warehousing, Star/Snowflake Models, Partitioning & Performance Optimizations, Power BI.

Cloud & Big Data (Azure): Azure Data Factory, Azure Databricks, Azure Data Lake Storage (Gen2), Delta Lake, Microsoft Fabric, Apache Spark.

CI/CD & DevOps: Azure DevOps, Git, Jenkins, CI/CD Pipelines

Domain Knowledge: Telecom OSS/BSS, Billing Systems, CDR/Usage Data, SLA & Performance Analytics

Experience

Netcracker Technologies

Dec 2022 - Nov 2025

Data Engineer

Hyderabad, India

VFHU (Vodafone Hungary)

- Processed large OSS/BSS datasets (customer, usage, billing, provisioning) using Azure Databricks (PySpark) and Synapse SQL to improve SLA visibility and data quality.
- Built and automated ADF pipelines to extract, clean, and transform telecom data from CRM, Billing, and Mediation systems into ADLS Gen2 and curated Synapse layers.
- Created data models and optimized transformation logic in Databricks/Delta to support high-volume telecom workloads.
- Performed cross-platform validation across ADLS, Synapse, and Databricks to reconcile mismatches between CRM, Billing, and Mediation systems.
- Diagnosed data defects by tracing lineage through ADF pipelines, notebooks, schema evolution, and ingestion logic, resulting in more reliable releases.
- Coordinated with CRM/Billing/Mediation teams to standardize ingestion, reconciliation, and reporting procedures on Azure.
- Supported migration cycles by validating ADF datasets, linked services, schema updates, and pipeline outputs prior to deployment.

TBO (Telenet Backoffice Automation)

- Extracted, cleaned, and standardized provisioning, customer, and usage datasets from OSS/BSS systems using Azure Data Factory and Databricks (PySpark) for automated downstream workflows.
- Built Power BI dashboards connected to Synapse/ADLS to track provisioning volumes, activation failures, SLA compliance, and operational KPIs.
- Performed root-cause analysis of provisioning failures by tracing data across CRM → Mediation → Billing using Databricks notebooks and targeted validation queries in Synapse SQL.
- Identified transformation and mapping defects in ADF/Databricks logic and proposed fixes that reduced repeated automation failures in provisioning flows.
- Collaborated with Billing, Provisioning, and CRM teams to diagnose recurring issues and implement stable Azure-based workflow improvements.
- Supported CI/CD deployments using Azure DevOps by validating ADF pipeline updates, linked services, and Databricks notebook changes.

Projects

End-to-End Azure Retail Data Engineering Pipeline (ADF • ADLS • Databricks • Power BI)

- Designed and implemented a full Azure data pipeline using ADF, ADLS Gen2, Databricks (PySpark), and Power BI for a retail analytics use case.
- Built Bronze-Silver-Gold layers in ADLS using the Medallion Architecture to separate raw ingestion, curated data, and aggregated business metrics.
- Developed ADF pipelines to ingest data from Azure SQL DB and an API (JSON), landing structured and semi-structured datasets into the Bronze layer.
- Performed data cleaning, type casting, deduplication, and multi-table joins in Databricks to produce a unified Silver dataset stored in Delta format.
- Engineered Gold-layer sales KPIs and created interactive Power BI visuals showing store performance, category trends, and monthly sales insights.

Education

Master of Computer Applications

JNTUH College of Engineering

Dec 2020 – Dec 2022

Hyderabad, Telangana

Certifications

Data Engineering Foundations Professional Certificate by Astronomer

Microsoft Azure AI Essentials Professional Certificate by Microsoft and LinkedIn