Sudheer Bandaru

Azure Data Engineer | Databricks

Phone: +91 8074941011 Email: sudheerbandaru34@gmail LinkedIn: sudheer-bandaru

Profile Summary

Result-driven Data Engineer with 5 years of experience in designing, building and optimizing scalable data platform solutions across real-time analytics, cloud technology, data integration and data warehousing environments. Experienced in end-to-end data engineering and passionate about integrating AI capabilities with modern data platforms to build intelligent, automated and business-driven solutions. Seeking opportunities to contribute to impactful data and AI initiatives while continuously expanding my technical expertise.

Technical Skills and Knowledge

- **Programming & Databases:** Python, SQL, PySpark, MySQL, SQL Server
- **Data Engineering & Cloud:** Azure Data Factory (ADF), ADLS Gen2, Azure Blob Storage, Azure Databricks, Azure Logic Apps, Azure Key Vault
- **Data Warehousing & Transformation:** Snowflake, Azure Synapse Analytics, ETL/ELT, Data Modelling, Data Migration, Data Pipelines
- **Orchestration, DevOps & Visualization:** Apache Airflow, Databricks Workflows, Git, GitHub, CI/CD, Power BI
- **AI & Engineering Concepts:** Machine Learning, Prompt Engineering, AI Agents - n8n, SDLC, Agile/Scrum

Professional Work Summary

Data Engineer |Senior Engineer| KPIT Technologies |Jul 2021 – Jun 2026

- Developed Python/PySpark processing logic for structured and semi-structured data sets with a focus on high-volume data sets
- ELT pipeline development (Azure Data Factory, Databricks, PySpark, Spark SQL) resulting in ~60% improvement in processing time
- Quality control and data validation across data sets to ensure consistency of data processed against source data sets
- Support of data migration, production support, production fixes, and pipeline optimizations in end-to-end data pipelines
- Developed end-to-end desktop applications supporting processing of high-volume data sets capturing data for business analytics and enabling reduced engineering efforts
- Process development for extraction, transformation, and loading of data sets in Snowflake data warehouse
- Performance tuning of data processing pipelines including optimization of joins, partitioning, and bucketing, data file formats and query rewrites

Projects Work Experience

Project 1 - HONDA - ADAS Data Cloud Data Platform OPERATIONAL DATA PIPELINE

Project Overview:
Built scalable data pipelines for processing ADAS and vehicle telemetry data, integrating automotive logs and enterprise data sources into a cloud-based data platform for validation and analytics.

Key Responsibilities & Achievements:

- Developed PySpark Databricks transformations for data cleansing, standardization, deduplication, validation, and enrichment.

- Implemented Bronze/Silver/Gold architecture, using ADLS for raw/processed data and Snowflake for curated analytical datasets.
- Implemented SCD Type 1 & Type 2 for dimension tables to maintain historical vehicle, test, configuration, and validation data.
- Configured 35+ ADF triggers and incremental pipelines aligned with processing SLAs.
- Implemented data quality and reconciliation checks for completeness, duplicates, schema mismatches, and timestamp consistency.

PROJECT 2: NESTLE USA - DIGITAL YWIN CLOUD MIGRATION (MATRIX PROJECT)

Project Overview:
Led cloud migration initiative modernizing sourcing, production, delivery, and demand generation processes for liquid beverages through strategic cross-functional data pipeline development, processing 500GB+ daily data volume and reducing operational costs by 25%.

Key Responsibilities & Achievements:

- Designed and implemented 40+ critical pipeline components independently, delivering 100% of milestones on schedule across 8-month project timeline
- Migrated 200+ tables and 2TB+ data from multiple sources (Snowflake, on-premises SQL Server, SharePoint, file systems) with zero data loss
- Loaded transformed data into Azure Synapse, achieving 100% data integrity, consistency, and alignment with business data models across 10+ subject areas
- Implemented Slowly Changing Dimension Type 1 and Type 2 for 20+ dimension tables, maintaining 5+ years of historical data
- Supported User Acceptance Testing (UAT) for 15+ business scenarios, achieving 98% test pass rate
- Applied 10+ production hotfixes with average resolution time of 2 hours, minimizing business impact to less than 0.1% downtime.

Data Science – Internship Project Experience

Intern | INNODATATICS | Oct 2020 – Jun 2021
Group Health Insurance Policy Analysis and Segmentation

- Cleaned, transformed and Analysed 1,179 group health insurance policies records across 20 attributes using Python, Pandas, including missing-value handling, outlier treatment, encoding and feature engineering.
- Performed data quality checks and exploratory analysis on premiums, coverage, claims ratios, group sizes, and other policy attributes.
- Applied K-Means and Hierarchical Clustering to segment insurance policies and evaluate meaningful data patterns and derived insights by comparing policy premiums, coverage features, claims ratios, group sizes, and network hospitals.

Education

Bachelor of Technology (B.Tech) – Electronics & Communication Engineering (ECE) | JNTUA-VITS | Andhra Pradesh | 2016–2020

Achievements

- **Manager Discretion spot Award (2025)** – Recognized for contributions to data pipeline development, ETL processes, and delivering project requirements on time.
- **The Comrades and High Flyer Award (2023 and 2024)** – Recognized for strong ownership, effective collaboration, and consistent delivery of key responsibilities.