Telugu Doddi Suresh

Contact: +91-9652358728 **Email:** telugusuresh.1998@gmail.com

Professional Summary

Results-driven Data Engineer and Data Analyst with over 4 years of experience in building scalable, cloud-native data solutions using Azure, Python, SQL, and PySpark. Specialized in designing robust ETL/ELT pipelines with ADF and Delta Lake and implementing Lakehouse architecture for efficient data processing. Proficient in working with structured and semi-structured data from sources like MongoDB, PostgreSQL, and Azure Blob Storage. Experienced in exploratory data analysis (EDA) to deliver actionable insights and support business decision-making. Strong background in data governance using Unity Catalog and Microsoft Purview, ensuring data quality, lineage, and compliance. Proven track record in performance optimization, cost reduction, and automation, with the ability to deliver end-to-end solutions for real-time dashboards, AI/ML use cases, and analytics in Agile/Scrum environments.

Core Competencies

- **Programming & Scripting:** Python (Pandas, NumPy), SQL
- **Big Data & Distributed Processing:** PySpark, Apache Spark, ETL/ELT Pipeline Development
- **Cloud & Data Platforms**: Azure Databricks, Azure Data Factory (ADF), Delta Lake
- **Databases:** PostgreSQL, MongoDB, NoSQL, Relational Databases
- **Data Modeling & Governance:** Dimensional Modeling, Data Quality and Integrity, Data Lineage, Unity Catalog, Microsoft Purview
- **Visualization & Reporting:** Power BI, Interactive Dashboards
- **Workflow Orchestration:** Azure Pipelines, Triggering & Monitoring Data Workflows
- **Professional Skills:** Feature Engineering, Performance Optimization, Modular Data Architecture, Agile/Scrum Collaboration

Professional Experience

Consultant-Python Data Engineer

Deloitte Consulting India Private Limited (USI), Hyderabad March 2026 – Present

Senior Data Engineer

LTIMindTree Limited, Hyderabad September 2025 – February 2026

Data Engineer

HashStack Solutions Private Limited (Formerly known as Zitisi Solutions LLP), Hyderabad October 2021 – August 2025

MAYAMAYA– AI-Powered Career Decision Platform

- Designed and optimized scalable ETL/ELT pipelines using **PySpark** and **Spark SQL** in **Azure Databricks** to process high-volume behavioural, psychometric, and transactional data from **MongoDB**, enabling advanced analytics and ML readiness.
- Leveraged **Apache Spark's distributed computing** to parallelize complex data transformations, drastically improving data processing efficiency and runtime scalability.

- Implemented **Delta Lake architecture** (Bronze → Silver → Gold layers) with schema evolution and time travel support for reliable, version-controlled batch and incremental data loads.
- Applied **Spark SQL optimizations** such as broadcast joins, predicate pushdown, and caching strategies, reducing compute cost and improving pipeline execution by 50%.
- Performed advanced data cleaning and transformation using **PySpark DataFrames**, **Pandas**, and **NumPy**, reducing preprocessing time by 50% and ensuring data integrity for ML models.
- Automated data ingestion and transformation workflows using **Azure Data Factory**, integrating with Spark jobs to enable continuous data refresh and minimize manual overhead.
- Designed **scalable data models** and **partitioning strategies** within Databricks and Delta Lake, resulting in a 20% improvement in query performance and a 50% reduction in operating costs.
- Developed and integrated **custom Spark-based data quality checks**, detecting anomalies early and reducing data transformation errors by over 70%.
- Migrated legacy pipelines from **PostgreSQL to Apache Spark in Databricks**, achieving improvements in scalability, maintainability, and throughput.
- Tuned **Spark configurations** (shuffle partitions, executor memory, parallelism) to optimize cluster performance, leading to a 30% improvement in SQL query speed and 50% reduction in data load latency.

CMA (Common Man Army)– Intelligent Image Data Extraction

- Developed a high-performance **Python automation solution** to extract structured data from **images embedded in PDF files** using OCR and computer vision techniques.
- Utilized **Tesseract OCR** with **OpenCV-based algorithms** such as **Template Matching**, **Edge Detection**, and **Feature Matching** for accurate image parsing.
- Applied **multi-threading** to enable parallel image processing, achieving a **75% boost in performance** and reducing total processing time significantly.
- Cleaned and normalized extracted data using **Regex**, **Pandas**, and custom logic to ensure high accuracy.
- Stored clean, structured output in **CSV format** and automated ingestion into a **PostgreSQL** database.
- Reduced manual data entry errors and significantly improved processing scalability for high-volume PDF workflows.

DHENUSYA ORGANICS– Billing Automation Application

- Developed Designed and implemented a lightweight **Python-based billing application** to automate invoice generation, inventory management, and data tracking.
- Enabled **CSV-based bulk data uploads**, dynamic invoice creation, and **date-based reminders** to streamline the billing cycle.
- Integrated **PostgreSQL** as a backend for secure and scalable data storage and retrieval.
- Leveraged **Pywhatkit** to facilitate **WhatsApp-based invoice delivery**, improving client communication and reducing operational delays.
- Reduced manual billing effort by over **60%** and enhanced overall accuracy and speed in financial workflows.
- Delivered a complete solution combining automation, data integrity, and real-time invoice sharing to support small business operations.

Key Achievements

- Designed and implemented scalable ETL pipelines using **Azure Databricks, PySpark, and Delta Lake**, improving data processing efficiency and enabling reliable integration with cloud storage and downstream analytics.
- Reduced operational costs by **50%** and enhanced system performance by **20%** through optimized data modeling, SQL query tuning, and schema restructuring in the **MayaMaya** project.

- Migrated data pipelines from legacy systems (PostgreSQL) to **Azure Databricks**, achieving **3x performance gains** and enabling seamless scalability.
- Ensured high data availability and reliability through proactive monitoring, failover strategies, and cloud-native pipeline orchestration using **Azure Data Factory**.
- Contributed to end-to-end **data governance**, lineage tracking, and performance optimization initiatives to support regulatory compliance and data quality across projects.
- Improved data quality by implementing robust validation frameworks and automated anomaly detection checks, leading to a **70% reduction in transformation errors**.
- Built an OCR-driven data extraction solution using **Python, Tesseract, OpenCV, and PyPDF2** for the **CMA** project, increasing document processing throughput by **75%** and significantly reducing manual data entry efforts.
- Automated key components of invoice management in the **Dhenusya Organics** billing application using Python, reducing manual effort by **60%** and improving billing accuracy.
- Integrated **WhatsApp-based invoice delivery** using Pywhatkit, improving client communication and operational responsiveness.

Technical Projects

1. **ETL Pipeline Automation – Azure Data Factory & Databricks**
   Designed and implemented scalable, production-grade ETL pipelines using **Azure Data Factory** for orchestrating data ingestion from **SQL databases** and **Azure Blob Storage**. Developed complex transformation logic in **Azure Databricks** using **PySpark**, including multi-level joins, window functions, aggregations, type casting, and schema validation. Stored curated data in **Delta Lake** format to enable high-performance querying, ACID compliance, and version control. Adopted modular and parameterized pipeline design, reducing manual intervention by **80%**, improving reusability, and enhancing data consistency across reporting layers.
2. **SQL Performance Optimization & Reporting Automation – Python & SQL**
   Optimized large-scale, complex **SQL queries** through index tuning, query refactoring, and execution plan analysis, resulting in a **50% improvement** in report and dashboard performance. Automated end-to-end reporting workflows using **Python scripts** and **SQL stored procedures** to schedule, generate, and distribute daily and weekly reports. This automation increased reporting accuracy, reduced manual errors, and freed up analyst time for higher-value tasks, significantly boosting overall team productivity.

Strengths

- Strong team collaboration with a focus on shared goals and knowledge sharing
- Proven problem-solving and critical thinking skills for resolving complex data challenges
- Excellent communication skills with adaptability across technical and non-technical teams
- Self-motivated, proactive, and resilient under pressure with a solution-oriented mindset

Education

Bachelor of Technology (2017 - 2021)-Malla Reddy Institute of Technology

Branch: **Computer Science and Engineering**

Declaration:

I hereby declare that the information provided above is true to the best of my knowledge and belief.

Place: Hyderabad (Telugu Doddi Suresh)