Shivaraj Galagali

Linkedin | Github Profile | shivarajsgalagali5@gmail.com | +919620030799

Professional Summary

Data Engineer with 2 Years 8 Months of experience in developing ETL pipelines, data warehousing solutions, and Python applications. I specialize in Azure services and big data technologies, with a strong focus on optimizing data workflows to enhance business intelligence and operational efficiency. Recognized for innovative problem-solving and effective collaboration.

Technical Skills

Data Engineering:	ETL/ELT, Data Modeling and Warehousing, SCD, Data Quality, Event-Driven Architecture.
Programming:	Python (DSA and Libraries), SQL, PySpark
Cloud and Integration:	Databricks, Azure (Synapse, Data Factory, blob, Logic Apps, Event Hubs, MSAL).
Database Technologies:	Azure SQL Server, Synapse Data Warehousing, MySQL, MongoDB
Libraries & Tools:	Pandas, Selenium, Power BI, Azure DevOps, Git

Work Experience

Happiest Minds Technologies (2022 August - Present)

Role: Data Engineer

Dover: Customer Data Platform (CDP), Supplier Data Platform (SDP)– Integrated Business Data (ERP) reporting for 20+ Independent Global Operating Units. **Master Data Management (MDM)**

- Developed 50+ Azure Data Factory pipelines for ERP data integration from multiple operating companies, centralizing customer, supplier, and transaction data for comprehensive reporting.
- Collaborated with ERP teams to validate data and create mapping documents; developed pipelines with complex queries, followed by rigorous testing and deployment to production.
- Conducted pre-escalation analysis to address data discrepancies (aging factors, receipts, missing invoices, spend mismatch), delivering actionable insights to the Business Operations Team.
- Delivered weekly post-production updates, implementing business changes (e.g., entity movements, managed accounts), minor modifications (e.g., column additions, BU mapping), and integration pipelines (e.g., JSON logs of customer inquiry data into CDP).
- Supported **MDM** through global entity ID validation, debugging elastic integration pipelines, and enhancing normalization logic for reliable data alignment across platforms.

Key Technologies: SQL Server, Azure Data Factory, Azure DevOps, Databricks, Python, Excel, PowerBI

TMF Group: Regulatory Bulletin - Regulatory Summaries for Business Enablement

- Transitioned to an event-driven architecture using Azure Synapse and Event Grid, replacing periodic database polling to automate batch processing. This shift resulted in a 35% reduction in processing time and doubled throughput.
- Created Flask applications for efficient scraping and data extraction, incorporating blob and database transactions, API authentication, object lifecycle management, logging, retries, and robust error handling, deployed with Gunicorn (WSGI server).

- Increased scraping success from 60% to 95% during POC employing advanced techniques, including anti-scraping circumvention, optimized request handling, session-based scraping and error recovery. This success was crucial for enabling further development.

Key Technologies: Python (Requests and Selenium Scraping), Flask, API development, Azure Data Factory, Azure Event Grid, Logic app.

TMF Group: Audit Logs - Enabling incremental Reporting of Kraios Platform user metrics.

- Implemented Databricks Autoloader for incremental ingestion of 40 GB daily AVRO log data, transitioning to a full PySpark ETL and reducing processing time by 20-40%, ensuring efficient log management of 8,000+ clients.

Key Technologies: Databricks, PySpark, Delta Table.

Education

National Institute of Engineering, Mysore   2018-2022
B.E in Computer Science (CGPA: 8.1)
Research Paper: Smart Street Lighting System – Adaptive and interactive system for energy conservation.

Awards and Recognitions

Learning Champion – Recognized for innovative Python solutions that enhanced efficiency in TMF's Regulatory Bulletin Project

Team Excellence Award - Awarded for critical contributions to TMF's Audit Logs and BDR projects

Certifications

- Microsoft Certified: Azure Data Engineer Associate (DP-203) - July 2024
- Data Engineering Training Certification by Stack Route – December 2022