Janani N B

jananibalki2001@gmail.com   —   8610962092
www.linkedin.com/in/janani-balakrishnan-6941bb20a

Professional Summary

Data Engineer with 3 years of experience building and optimizing data pipelines using **Databricks**, **PySpark**, **Python**, and **SQL**. Developed scalable ETL workflows and notebooks within Databricks, leveraging **Azure Cloud** services including **Data Factory**, **Azure Functions**, and **Blob Storage**. Hands-on experience with integrating **OpenAI technologies** and implementing **RAG (Retrieval-Augmented Generation)** models to enhance intelligent data workflows. Proven ability to reduce runtime, improve code performance, and deliver production-ready solutions.

Technical Skills

Languages:	Python, PySpark, Spark SQL
Tools & Technologies:	Databricks, Azure Data Factory, Azure Functions, Azure DevOps, Blob Storage
Concepts:	ETL/ELT, Data Modeling, Data Warehousing, Data Lake Architecture

Professional Experience

Senior Software Engineer in Data Engineering, Tiger Analytics   **July 2023 – Present:**

- Developed an automated pipeline that integrated with DealCloud to streamline meeting workflows,collected upcoming meeting data, sent personalized email templates to attendees via Graph API 1 hour prior, and post-meeting, processed email responses as summaries, key points, and next steps, which were automatically updated in DealCloud. This solution improved meeting preparedness and eliminated manual data entry.
- Engineered an end-to-end data ingestion and transformation pipeline to extract content from scanned and non-scanned PDFs, applied preprocessing and chunking logic using LLMs, and stored resulting embeddings in Databricks Vector DB as part of a scalable, domain-specific RAG (Retrieval-Augmented Generation) architecture.
- Developed an agentic workflow that executes tasks based on user queries and selections from a frontend app, featuring multiple agents including a master agent, task-specific agents for retrieving data from vector databases and Databricks SQL tables, and a validation agent to ensure data accuracy.

Software Engineer in Data Engineering, Tiger Analytics   **May 2022 – July 2023:**

- Handled daily data engineering tasks including debugging pipelines, migrating pandas to PySpark, and developing robust ETL notebooks; optimized Spark workflows to reduce runtime by **83%** (2 hrs to 20 mins) and cut compute costs by over **80%**.
- Automated an event-driven pipeline using Azure Functions and Databricks Jobs to reduce idle compute time, saving over **75%** in monthly costs.
- Developed reusable data quality validation framework in PySpark to enforce schema checks, null constraints, and anomaly detection across ETL pipelines, improving reliability and reducing data issues in production.
- Developed a generative AI-based utility that parsed and summarized Databricks notebooks and workflows by extracting code, markdown, and metadata, reducing documentation effort and improving onboarding efficiency for new team members.

Education

Bachelor of Engineering in Electrical and Electronics Engineering

Sri Krishna College of Technology   Graduated: 2022

Certifications

- Databricks Certified Data Engineer Associate
- Microsoft Certified Azure Data Fundamentals