

Bharath Raavi

[Portfolio](#) | [LinkedIn](#) | [git hub](#) | braavi24@gamil.com

Summary:

Experienced Data Engineer with 4+ years of hands-on work in building and optimizing data pipelines using **Databricks** (Spark SQL/Scala). Skilled in ETL development, data modeling, Python scripting, and maintaining secure, well-documented data workflows for analytics and reporting.

Professional Experience:

Jhonson & Jhonson, New Jersey

April 2025 – Present

Data Engineer

- Designed and deployed robust data pipelines using Azure Data Factory, Databricks, and Delta Lake, powering daily data ingestion and transformation workflows.
- Engineered complex business logic with Spark SQL, Scala, and Python, enabling accurate, production-grade datasets for reporting and analytics.
- Implemented MERGE (UPSERT) logic to manage incremental and historical loads, enhancing data consistency and reliability.
- Proactively monitored and troubleshoot pipeline failures in ADF and Databricks, ensuring high availability and low-latency data flow.
- Authored detailed documentation on Confluence and partnered with analysts and architects to deliver secure, scalable data solutions.

Saint peter's University, New Jersey

Sep 2024 – Feb 2025

TA & Applied Research

- Led SQL tutoring sessions and guided students through real-world data challenges, improving query proficiency and engagement.
- Conducted applied research using the NHANES healthcare dataset (5,000+ records), developing XGBoost and Random Forest models to predict anemia with 92% accuracy.
- Evaluated models using precision (89%), recall (91%), and F1-score (90%), generating actionable insights for early medical intervention.
- Supported curriculum delivery and hands-on labs, bridging theory and practice for data science learners.

Wipro, India

Mar 2020 – Jul 2023

Data Engineer

- Built and maintained scalable ETL pipelines using SQL, Snowflake, Python, and AWS S3 to enable efficient, cloud-based data processing.
- Developed Snowflake views, stored procedures, and automated workflows with Apache Airflow, reducing manual overhead and ensuring data reliability.
- Performed end-to-end data quality checks and collaborated on query optimization, access control (RBAC), and IAM policy management.

Independent Projects (Recent):

- Spotify Music Analytics Pipeline – Azure Cloud (Capstone Project) - [Git Link](#)**
Built an end-to-end scalable ETL pipeline on Azure to process Spotify streaming data with 60% less manual effort.
- AI-Powered ETL Code Generator Using GPT-2 and Flask - [Git Link](#)**
Developed a web app using GPT-2 that generates Python ETL scripts from natural language inputs instantly.

Skills: Python, SQL, Scala, Spark, Azure Data Factory, Azure Databricks, AWS, Databricks, Delta Lake, Medallion Architecture, Snowflake, Hugging Face Transformers, PyTorch, ollama, Flask, HTML/CSS, SQL query optimization, Git, GitHub, bit bucket, confluence, Jira.

Education:

Saint peter's university, New Jersey

Aug 2023 – Feb 2025

MS in Business Analytics

Certifications:

[Azure AI Engineer Associate](#), [Azure Data Engineer Associate](#), [Data Engineer Boot Camp by Data Expert.io](#)