

Bharath Raavi

[Portfolio](#) || [Linkedin](#) || [GitHub](#) || braavi24@gmail.com || Englishtown, NJ 07726

Professional experience

Data Analyst || Wipro Pvt Ltd || November 2020- July 2023

- Maintained and optimized data pipelines using SQL, **Python**, **Apache Airflow**, and Spark, ensuring efficient data ingestion, transformation, and loading (ETL).
- Implemented transformations in Spark to process large-scale datasets, improving pipeline performance by **40%** for compliance reporting.
- Automated ETL workflows using Apache Airflow and SQL, reducing manual **effort by 60%** and ensuring real-time data availability.
- Developed and optimized SQL stored procedures for complex data transformations, business logic implementation, and performance tuning.
- Conducted pipeline testing, validating data accuracy, transformations, and performance to ensure robust data flows.
- Implemented data quality rules in SQL and Python, performing anomaly detection and error handling to enhance data reliability.
- Optimized complex SQL queries in **Azure Synapse & SQL Server**, improving query performance by 30% through indexing, partitioning, and query tuning.
- Developed Python-based validation scripts to troubleshoot data pipeline failures and resolve ingestion issues efficiently.
- Collaborated with business and engineering teams, gathering technical requirements and implementing data-driven solutions aligned with compliance needs.
- Ensured data security, governance, and compliance, managing audit tracking and enforcing access controls across teams.

Teaching Assistant-Volunteer | Saint Peter's University | September 2024 - November 2024

- Led sessions to guide students in improving their SQL query skills.
- Provided hands-on support and facilitated lessons, contributing to higher student engagement and improved project outcomes.

Applied Research | Saint Peter's University | November 2024 – February 2025

- Built and optimized XGBoost and Random Forest models for anemia prediction, achieving 92% accuracy on the NHANES dataset with over 5,000 participants.
- Implemented data preprocessing techniques, reducing missing data impact by 15%, and applied feature selection to identify key health indicators.
- Evaluated models using precision (89%), recall (91%), and F1-score (90%), enabling data-driven healthcare insights for early anemia detection.

Independent Projects (Recent)

Spotify Music Analytics Pipeline in Azure Cloud

- Built an ETL pipeline to extract and process **Spotify API data using Python, SQL**, and Azure data factory, reducing manual effort by 60%.
- Optimized data processing in Azure Databricks and Synapse, improving query speed by 30%. Stored data according to **Medallion architecture**
- Ensured 99% data accuracy with **SQL-based** validation and quality checks.

AI-Powered ETL Code Generator Using GPT-2 and Flask

- Developed a Flask web application that leverages a pre-trained **GPT-2 model to generate** ETL (Extract, Transform, Load) Python scripts from user-provided descriptions.
- Integrated **Hugging Face Transformers and PyTorch**, enabling natural language processing to automate ETL code generation with minimal setup. Designed a user-friendly interface for seamless interaction.

Education

Saint Peter's University

MS in Business Analytics – GPA: 3.89

Sep2023- Feb 2025

Jersey City, New jersey

Skills

Programming & Databases: python, PySpark, Transact-SQL, SQL (NoSQL, RDBMS, Redis, oracle, MongoDB, My SQL, PL SQL, MS SQL)

Big Data & AI/ML: Hadoop, Scala, Kafka, Spark, Databricks, Delta Lake, Air flow, Automic, spark streaming, ETL, AI/ML, Gen AI

Cloud: Azure (ADF, Synapse Analytics, Synapse SQL Server, Databricks, Dataflows, Azure Event Hub, Azure AI/ML), GCP

Project Management & Communication: Agile Methodologies, Cross functional teams, Jira, Risk managent, Business Communication

Certifications

Microsoft Azure: [Certified Azure AI Engineer Associate](#), [Certified Azure Data Engineer Associate](#)

other: [SQL\(Advanced\)](#) , [SQL\(Intermediate\)](#), [PWC Power BI Job Simulation](#), [Accenture Data Analytics and Visualization](#)