Bharath Raavi

Portfolio || **LinkedIn** || **Git Hub** || braavi24@gmail.com

Professional Summary

I am a Data Engineer with 4 years of experience with strong skills in Snowflake, SQL, and Azure. Built and managed data pipelines, handled large datasets, and worked on data warehouses for reporting and analytics. I am good at writing complex SQL queries, creating ETL workflows, and improving data quality. Worked with tools like Azure Data Factory, Databricks, and Synapse. Focused on building reliable, clean, and secure data systems.

Professional Experience

Wipro Pvt Ltd, India.
Data Engineer || Mar 2020 – Jul 2023

- Help build and optimize ETL pipelines using **SQL**, **Snowflake**, and **Python**, with some use of **AWS S3** for storing and loading raw data.
- I worked on writing **SQL queries** and creating views in Snowflake to clean, transform, and prepare data for analysis and reporting.
- Assisted in automating data workflows with **Airflow**, making routine tasks more efficient and reducing manual effort.
- Contributed to writing and improving **Snowflake stored procedures** and SQL logic used in day-to-day business operations.
- Took part in testing and validating data pipelines to make sure the data was accurate, consistent, and ready for use.
- Used **SQL scripts** to check for missing values, duplicates, and other quality issues in the data; also experimented with Python for basic checks.
- Learned how to review and understand query execution plans in Snowflake and made small changes to improve performance.
- Worked with senior developers to troubleshoot issues during data loading and transformation, especially when something failed or looked off.
- Attended team meetings and worked with analysts and engineers to understand their data needs and support secure access to the right datasets.
- Gained practical experience setting up user roles and permissions in **Snowflake** and learned the basics of access control using **S3** and IAM.

Saint Peter's University, Jersey City.
Teaching Assistant || September 2024 – Nov 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 || Nov 2024 – Feb 2025

- Built and optimized XGBoost and Random Forest models from scratch to predict anemia using the NHANES dataset (5,000+ records), achieving 92% accuracy.
- Applied data preprocessing techniques to handle missing values, reducing their impact by 15%, and performed feature selection to identify key health indicators.
- Evaluated model performance using precision (89%), recall (91%), and F1-score (90%), providing actionable healthcare insights for early anemia detection.

Independent Projects (Recent)

Spotify Music Analytics Pipeline – Azure Cloud (Capstone Project) - Git Hub Link
Tech Stack: Python, Azure Data Factory, Azure Databricks, Azure Synapse, SQL, Medallion Architecture

- Designed and implemented a complete **ETL pipeline** to extract, transform, and load Spotify user data using **Spotify API**, **Python scripts**, and **Azure Data Factory**, reducing manual data collection by **60%**.
- Processed and transformed large volumes of streaming data into **Azure Databricks**, then loaded clean data into **Azure Synapse Analytics** for downstream analytics and reporting.
- Followed **Medallion Architecture (Bronze, Silver, Gold layers)** to organize raw, refined, and curated data for better scalability and governance.
- Applied **SQL-based validation checks** and transformation logic to ensure **99% data accuracy**, enabling reliable insights into user behavior and artist popularity.
- Focused on **query optimization**, improving performance by **30%** through indexing and partitioning strategies.

AI-Powered ETL Code Generator Using GPT-2 and Flask - Git Hub link
Tech Stack: Python, Flask, Hugging Face Transformers, PyTorch, HTML/CSS

- Created a web-based tool that automatically generates **Python ETL scripts** from plain text descriptions using a fine-tuned **GPT-2 model**.
- Used **Hugging Face Transformers** and **PyTorch** to enable natural language understanding and translation into functional ETL code.
- Designed and deployed a **Flask application** with an intuitive user interface that allows users to input job requirements and instantly receive code outputs.

Skills Set

Programming & Databases: Python, SQL, PySpark, Transact-SQL, PL/SQL, MySQL, PostgreSQL, MongoDB, Oracle, RDBMS, JSON, REST API
Big Data & Data Engineering: Spark, Kafka, Hadoop, Delta Lake, Airflow, ETL/ELT Pipelines, Snowflake (Streams, Tasks, Snowpipe), Data Modeling (Star/Snowflake), SCD, Data Quality, Data Validation, Medallion Architecture, Batch & Stream Processing
Cloud: AWS (S3, Glue, Lambda, Redshift, RDS, Athena, Step Functions, IAM, CloudWatch**), Azure (**Data Factory (ADF), Synapse Analytics, Blob Storage, Azure Databricks, Azure SQL), **Snowflake (**SQL, UDFs, Stored Procedures, Streams & Tasks, Clustering, Time Travel)
Tools & Frameworks: Git, GitHub Actions, Jupyter, VS Code, Flask, MLflow, Hugging Face, PyTorch
Project Management & Communication: Agile Methodologies, Jira, Confluence, Cross-functional Teams, Risk Management, Documentation, Business Communication.

Education

Saint Peter's University — Sep 2023- Feb 2025
MS in Business Analytics – GPA 3.87 — Jersey City, New Jersey

Professional certifications

Cloud: Azure Data Engineer Associate, Azure AI Engineer Associate
Boot Camp: 5 weeks Data Engineer boot camp by DATA Expert.IO
Virtual Internship from forage: Accenture Data Analytics, PWC Data Analytics
SQL: Hacker rank Advanced SQL certification, Hacker rank Intermediate SQL certification