Praful Nair

(514) 573-7144 | prafulworkspace@gmail.com | LinkedIn | GitHub| Portfolio

Education

Concordia University, Montreal Sept. 2022 – June 2024
Masters in Applied Computer Science | ***GPA**: 3.65/4.3* *Montreal, QC, Canada*
- **Courses**: Data Engineering, Distributed Systems, Computer Vision, Algorithm Design Techniques

University of Mumbai, Maharashtra, India Aug. 2016 – Jan. 2021
Bachelors in Computer Engineering | ***GPA**: 8.42/10* *Mumbai, India*

Technical Skills

Languages: Python, SQL, Java, JavaScript, C **Libraries/Tools**: pandas, NumPy, matplotlib, seaborn, Tableau, Flask, Git, GitHub Actions, Docker **Databases**: MySQL, PostgreSQL, BigQuery **Cloud**: GCP (Cloud Functions, BigQuery), AWS (EC2, S3) **Other**: Data Cleaning, Exploratory Data Analysis (EDA), Data Wrangling, Visualization

Experience

Developer Intern Sept. 2024 – Dec. 2024
Valsoft Corporation *Montreal, QC, Canada*
- Developed a MapReduce script and API-based service to sync invoice data from NetSuite ERP to a third-party system, ensuring high reliability across financial datasets
- Used log analysis and retry logic to improve visibility and reliability in data transfer pipelines across accounting systems
- Built an internal platform for feedback management using Flask and PostgreSQL, with dashboards to track data-driven insights from cross-team feature suggestions

Graduate Teaching Assistant Sept. 2023 – May 2024
Concordia University *Montreal, QC, Canada*
- Assisted students in understanding Agile development and managing structured software processes with a strong emphasis on data accuracy and traceability
- Reviewed data-related project submissions and guided students on maintaining clean datasets, consistency, and logical analysis in software projects

Projects

U.S. Flight Data Analysis | *Python, pandas, EDA, Data Cleaning* May 2024
- Analyzed U.S. 2021 flight data to extract insights on canceled and diverted flights, average airtime, and missing departure time data using Python and pandas
- Identified trends in flight cancellations during September 2021 and quantified diverted flights for targeted November windows
- Calculated average airtime for specific routes and handled missing data effectively to clean noisy CSV datasets
- Used Jupyter/Colab to present modular notebooks on sub-tasks: cancellations, diversions, airtime, and missing data detection

GCP Data Streaming ETL Pipeline | *GCP, BigQuery, Cloud Functions* Dec. 2022
- Built a cloud-native ETL pipeline using GCP Cloud Functions to process Amazon book review data from Cloud Storage to BigQuery
- Optimized payload sizes and trigger timings, improving pipeline efficiency by 40%

QueryQantify - RA Query Estimator | *Python, SQL Estimation, Query Optimizer* Jan. 2024
- Built a Python-based system to estimate output sizes of RA queries using statistical metadata
- Simulated database optimizers to compute intermediate result sizes, aiding in query planning and cost estimation

Network Flow Optimization | *Graphs, Algorithms, Python* Dec. 2023
- Implemented Ford-Fulkerson with different path strategies; benchmarked on simulated traffic networks (200–500 nodes)
- Analyzed performance and capacity bottlenecks across randomized datasets

Certifications

Google Data Analytics Professional Certificate — Coursera