MICHAEL SHEROUBI

msheroubi@gmail.com | msheroubi.gtihub.io | GitHub/LinkedIn

BSc. Computer Science | Security+

EXPERIENCE

Data Analyst & Support. *Clearmind International*   2020 - Present
- Developed an ETL solution using Python to migrate over 7 years of historical data; from legacy systems & csv files to a cloud-based solution.
- Designed OLTP database to handle simultaneous transactions by 20+ instructors while maintaining ACID compliance & data integrity.
- Normalized transcript data to improve query performance, reducing query run-time by over 75% and increasing accuracy of search results.
- Optimized registration process by modelling its flow of data and removing waste, trimming operational costs by 23%.
- Ensured enterprise-wide compliance with federal & international data protection regulations.

Data Analyst. *Asurion LLC*   2023 (6 months)
- Fine-tuned a large language model to preemptively detect prompting ineligible products.
- Developed a clustering algorithm using NLP and cosine similarity, reducing the workload of the Digital Mapping team by an average of 88% per batch.
- Designed a suite of testing scripts to help automate QA by cross referencing data across multiple batches.
- Contributed to a collaborative team project focused on developing a machine learning classifier.

Data Warehouse (Capstone), *Trellis Social Enterprise*   2019 - 2020
Engineer / Team Lead
- Designed, developed, and implemented a data warehouse solution using AWS Redshift, combining data from 5 disparate sources to prepare them for analysis.
- Developed an ETL solution to migrate data in batches from mongo DB to the data warehouse using Python scripts & task scheduler.
- Created and executed automated test plans, cases, and scripts to validate data integrity, and identify software issues & their root causes.
- Designed and optimized complex SQL procedures called by the company's internal dashboard.

SKILLS

Tech Stack: Python (pandas, NumPy, TensorFlow, Beautiful Soup, Matplotlib, PIL, opencv2), huggingFace, Langchain, pinecone
Java, SQL, Excel, AWS Redshift, Linux
Theory: Date Analysis, Database Design, ETL, Graph Theory, Natural Language Processing, LLMs

EDUCATION

University of British Columbia
BSc. Computer Science
- Honours Program | Minor in Mathematics / Data Science* | 3.90 GPA on 4.33 scale

Certifications: CompTIA Security+, Neo4j Certified Professional, Google Data Analytics

PROJECTS: ACADEMIC & EXTRA-CURRICULAR

Honours Thesis, *University of British Columbia*
Benchmarking performance for Neo4j in a social media application
- Developed automated test suites using Python to benchmark & document a Neo4j database's performance compared to a database using SQL Server.
- Generated and processed over 5 million rows of interconnected test data using Python.
- Created matching queries in SQL and Cypher to derive insights and answer questions from the social network.

OTHER

English: Native/Fluent   **Arabic**: Native/Fluent   **French**: Limited Working