Abigail Nadua

Trece Martires City, Cavite, 4109 – Philippines

☐ +63 976 564 7683   •   ✉ abinadua07@gmail.com

in linkedin.com/in/abigail-nadua/   •   ○ github.com/acnadua

Professional Summary

Results-driven Data Engineer with experience in designing and optimizing scalable ETL/ELT pipelines, automating data collection, and managing semi-structured data. Proficient in Python, SQL, Apache Airflow, and containerization. Passionate about leveraging cloud technologies (AWS) and Big Data tools (PySpark) to build robust data infrastructure. Proven track record of improving data throughput by 45% and pipeline success rates by 20% in a production environment.

Technical Skills

- **Languages:** Python, SQL, Dart, R, C, JavaScript/TypeScript, HTML/CSS
- **Data Engineering & Orchestration:** Apache Airflow, PySpark, Docker, Redis, Pandas, NumPy
- **Cloud & Databases:** AWS (S3), PostgreSQL, MongoDB, Firebase
- **Machine Learning & AI:** TensorFlow/Keras, Scikit-learn, LLM API Integration
- **Development Tools:** Git, Fast API, React, Node.js, Flutter

Work Experience

Media Meter Inc.   **Diliman, Quezon City**
Data Engineer   *July 2025 – Present*
- **Pipeline Optimization:** Optimized new and existing ETL/ELT pipelines for structured and semi-structured data, improving pipeline success rate by 20% and increasing overall data throughput by 45%.
- **Scalable Architecture:** Designed and implemented scalable data workflows using **Python**, **Redis**, **MongoDB**, and **Apache Airflow**, applying pub/sub architectures for efficient data processing.
- **Automated Data Ingestion:** Developed automated browser-based data extraction pipelines using **Playwright**, **Camoufox**, and **Drissionpage**, replacing manual tasks and enabling real-time ingestion with minimal intervention.
- **Observability & Monitoring:** Created observability dashboards and integrated logging mechanisms to monitor pipeline execution, enabling rapid troubleshooting and ensuring operational reliability.
- **Collaboration & Governance:** Collaborated with stakeholders to translate business requirements into technical solutions, documenting data architecture and operational procedures for internal governance.

University of the Philippines Los Baños   **Laguna**
Computer Science Instructor   *August 2023 – June 2025*
- Facilitated laboratory instruction for courses in Discrete Mathematics, Computational Methods, Automata Theory, and Mobile Development.
- Constructed tailored programming activities and resources, strengthening computational foundations and enhancing student comprehension of complex algorithms.
- Developed Android applications using Flutter and Firebase as teaching aids to demonstrate real-world industry practices.
- **Research Contribution:** Contributed to Natural Language Processing (NLP) research by curating and pre-processing large-scale datasets, utilizing keyword filtering and domain-specific lexicon development to isolate data relevant to public discourse on social issues. *(See Key Projects: NLP Data Curation)*

MyCode Corp. | **Quezon City**

Curriculum Developer | *March 2023 – August 2023*

- Conducted research on curriculum standards to improve Python and Web Development programs.
- Revamped curriculum resources, resulting in a 25% increase in resource accessibility and usability.

Key Projects

Personal Project

End-to-End Reddit Data Pipeline with LLM Analysis | *In Progress*

- **Orchestration & Containerization:** Architecting a fully containerized ETL pipeline using **Docker** and **Apache Airflow** to orchestrate daily data workflows.
- **Data Lake & Processing:** Implementing a custom ingestion layer using **Playwright** for subreddit discovery and **Python Requests** for JSON extraction, storing raw data in **AWS S3** and transforming it via **PySpark**.
- **LLM Integration:** Utilizing LLM APIs to perform sentiment classification on processed text data, removing the need for manual model training and maintenance.
- **Warehousing:** Loading transformed and enriched data into **PostgreSQL** for downstream analysis and reporting.

Personal Project

NASA Near-Earth Object (NEO) ETL Pipeline | *2025*

- **API Ingestion Strategy:** Developed a Python-based extraction pipeline to interact with the **NASA NeoWs API**, implementing robust error handling and pagination logic to fetch daily asteroid data.
- **Data Transformation & Logic:** Wrote transformation scripts to flatten complex nested JSON responses, standardize measurement units, and calculate hazard metrics based on object diameter and miss distance.
- **Data Modeling:** Designed a relational database schema in **PostgreSQL** to store historical approach data, optimizing for time-series queries to track potentially hazardous objects over time.

Academic Research Contribution

NLP Data Curation for Gender Bias Research | *2024*

- **Large-Scale Data Curation:** Curated and managed data sourced from the **NOW Corpus (English-Corpora.org)**, processing a dataset estimated to contain billions of tokens.
- **Targeted Preprocessing:** Utilized **Python**, **Pandas**, and **NLTK** to implement specialized cleaning, keyword filtering, and domain-specific lexicon development.
- **Research Focus:** Successfully isolated data patterns concerning the use of gendered pronouns (e.g., *he/she*) in connection with professions (e.g., *doctor, nurse*), enabling analysis of subtle gender biases in public language.

Academic Project

Cliquest: Cross-Platform Travel Coordination App | *2023*

- Designed a mobile app for real-time trip coordination using **Flutter**, **Node.js**, and **PostgreSQL**.
- Implemented WebSocket-based notifications for multi-user trip management and real-time communication.

Undergraduate Thesis

Sentiment Analysis of Tweets | *2022*

- **Deep Learning Models:** Developed and optimized deep learning architectures including **CNN** and **Bi-LSTM** with **Attention Mechanisms** using **TensorFlow/Keras** to classify social media sentiment.
- **Feature Engineering:** Engineered a text preprocessing pipeline using **Pandas** and **NumPy** to clean raw tweet data and implemented **GloVe embeddings** for semantic representation.
- **Analytical Impact:** Evaluated model performance to derive actionable insights from social media trends, providing the foundational research for the automated pipeline project currently in progress.

Education

University of the Philippines Los Baños | **Laguna**

Bachelor of Science in Computer Science, Magna Cum Laude | *Aug 2018 – Jan 2023*