

Eugene Thomas

Data Engineer

Suwanee, GA • etho4@uab.edu • 678-656-7155

Profile Summary

- **Data Engineer** with a robust academic foundation in **Computer Science** from Per Scholas, coupled with hands-on experience in managing large datasets and enhancing data-driven decision-making processes.
- Technically Proficient in a variety of programming languages including **Python**, **JavaScript**, and **SQL**. Well-versed in data management systems like **MySQL**, **Nifi** and **Airflow**, and skilled in employing data visualization tools such as Power **BI**, **Tableau**, and **Google Looker Studio** to transform complex data sets into actionable insights.
- Experienced in Data Transformation and Optimization through internships, managing over 800,000 natural product records from global sources, utilizing technologies like **AWS Redshift**, **AWS cloud services**, and **AWS EC2**, to streamline data processes and increase system efficiency and accuracy.
- Skilled Communicator and **Agile Practitioner**, effectively coordinating with **cross-functional teams** to deliver projects under tight deadlines. Demonstrates the ability to translate technical details into clear, strategic narratives.
- Committed to **Data Security** and **Data Quality**, implementing robust security measures including encryption and multi-factor authentication to meet industry standards and reduce vulnerabilities.
- Active Contributor to **Non-Profit Organizations**, such as the American Red Cross and the Medical Reserve Corps, where the candidate applied data engineering skills to enhance operational efficiency and emergency responsiveness.

Technical Skills

Programming Languages: Python, Javascript, SQL, HTML, CSS, Pyspark

Data Management: MySQL, Data Factory, Airflow, Luigi

Data Visualization: Power BI, Tableau, Google Looker Studio, AWS

Software Tools: NetSuite, Git, Github,

Education

University of Alabama at Birmingham

Birmingham, AL

Bachelor's degree in Psychology

Aug. 2021 - May 2025

Per Scholas

Birmingham, AL

Computer Software Engineering

Oct. 2024 - Jan. 2025

Work Experience

Ainnocence

Birmingham, AL

Data Engineer

Dec. 2024 - Present

- Developed and implemented a predictive bioinformatics CICD pipeline using Python to estimate organ-specific biological ages from multi-dimensional proteomic datasets (7,000+ features).
- Streamlined preprocessing workflows by resolving data inconsistencies, integrating demographic parameters, and automating pipelines, reducing manual input by 30%.
- Conducted statistical evaluations (e.g., z-scores, gap analyses) of treatment effects to rank organ changes and identify top gene candidates driving therapeutic insights.

- Created visualization dashboards, and analyzed datasets with Matplotlib, Numpy, Pandas and Plotly, delivering actionable data stories through comparative analyses of predicted and actual organ ages.

Built a localized Flask-based front-end portal enabling real-time patient data uploads and automated visual reporting for client use.

- Delivered high-priority project milestones under tight timelines, combining technical expertise with effective stakeholder communication to meet all client objectives.
- Trained and deployed advanced machine learning models leveraging **Vertex AI, Google Compute Engine, and Google Cloud Storage**, optimizing organ age prediction models by integrating distributed computing and efficient data management workflows for high-dimensional data, reducing training time by over 40%, increasing the accuracy of the models by over 35%, and delivering high quality training data and efficient organ prediction models.

Freelancer.com

Remote

Data Engineer (Contract)

Aug. 2024 - Dec. 2024

- Developed and executed a **Large Language Model (LLM)** project, focusing on data science topics by utilizing **TensorFlow**, and **PyTorch**, incorporating creative prompt engineering, efficient neural network optimization and ensuring that the model met the specified standards and objectives of clients. Furthermore a substantial dataset was created to train the model for the specific use case needed.
- Conducted intake meetings with customers to understand and analyze their business goals and align recommendations to their overarching strategy. Conveyed **challenges** and **opportunities** in a clear and concise way, adapting the message to each client's specific technical literacy.
- Evaluated **technical requirements** and researched adequate technologies, analyzing benefits and drawbacks in terms of scalability and sustainability, and aiming at avoiding future technical debt.
- Maintained clear and effective communication throughout the project lifecycle, ensuring alignment with project goals and facilitating seamless collaboration among team members.

University of Alabama at Birmingham

Birmingham, AL

Data Engineer

May 2024 - Jun 2024

- Created, cleaned, and processed a database of natural products using batch processing on historical datasets, including antibacterial, antifungal, and general wellness products, for a team of AI researchers and data scientists, playing a pivotal role during each phase of the **data life cycle**.
- Gathered approx. **20 terabytes** of data on natural products from organic chemical archives in Africa, Taiwan, US, Canada, and China, utilizing a **batch processing data ingestion framework** via **Microsoft Azure** to ensure data integrity.
- Transformed the database containing **800,000** products using **SQL**, successfully achieving an increase in system efficiency.
- Designed and implemented a comprehensive database from scratch, incorporating normalization and deduplication techniques alongside **ETL** pipelines to optimize data processing and storage efficiency.
- Implemented a data monitoring and optimization strategy to discover organic chemical remedies for illnesses through machine learning techniques, employing real-time data processing and **Apache Spark**, enhancing the predictive precision.
- Transformed raw data into actionable insights using **Python** and **SQL**, and employed data visualization in presentations to improve business decision-making.
- Enforced measures to guarantee security and compliance by integrating **multi-factor authentication** and utilizing **Zero Trust architecture**, thus achieving adherence to industry standards and reducing potential security breaches.
- Collaborated within a cross-functional team as well as the **CEO**, employing **Agile** practices such as daily stand-up meetings and weekly sprints, to effectively deliver software under tight deadlines.
- **Presented** solutions to data scientists, machine learning engineers, and the CEO, utilizing storytelling techniques to convert intricate data workflows into clear, comprehensible narratives that ensured understanding and alignment across specialized groups within the organization.

American Red Cross | Atlanta

Atlanta, GA

Data Coordination Volunteer

Jan. 2020 - Dec. 2021

- Participated in the technical effort by managing and updating the American Red Cross database to maintain accurate records of donor blood types and health conditions, delivering clean and scalable solutions.
- Facilitated communication between donors and managers through strategic scheduling of blood donation appointments and offered timely support during critical donation processes.
- Explored potential blood donors using **SQL**, and monitored **PowerBI** for **real-time analytics**, achieving a match rate of **82%** within the databases.
- Convened multiple sessions with the **management team**, including conference calls and face-to-face discussions, employing collaborative strategies to evaluate and advance the organization's progress.

East Metro Health District Medical Reserve Corps, Inc. (MRCGEM)

Atlanta, GA

Emergency Preparedness Volunteer

Jan. 2019 - Dec. 2021

- Organized various datasets and databases with **MySQL**, employing **data cleansing techniques such as deduplication and quicksort** to enhance the emergency verification process.
- Administered **emergency relief** and conducted structured training sessions on preparedness, employing best practices and standardized protocols to ensure readiness for unforeseen circumstances.
- Received **recognition** for providing critical emergency assistance during the pandemic by swiftly deploying robust technical solutions and leveraging advanced data analytics to enhance system responsiveness and reliability.

Certifications

Goldman Sachs - Software Engineering	Aug. 2024
MongoDB & Node.js	Aug. 2024
Verizon - Cloud Platform	Aug. 2024
Introduction: Data Engineering	Jun. 2023
MySQL assessment Hands-On	May. 2023
British Airways - Data Science	Jan. 2025