Wesley Lau

(571) 354-9369 • [wesleylau.wcl@gmail.com](mailto:wesleylau.wcl@gmail.com) • [LinkedIn](https://www.linkedin.com/in/wesley-lau-34a8b374/)

**Professional Summary**

Data Engineer with 6+ years of experience building scalable ETL pipelines and data warehouses using AWS, Python (PySpark), and SQL. Proven expertise in process automation, cost reduction, and leveraging big data technologies like Neo4j to support advanced analytics.

**Professional Experience**

**Data Engineer** | National Cancer Institute **November** 2022 – January 2025

* Led the integration of two external systems into the existing data architecture, expand the existing Neo4j Graph data model, improving data accessibility and scalability.
* Developed and executed a data migration strategy to transfer 3TB of data into Neo4j, utilizing Python-based ETL scripts with ***Pandas*** and ***NumPy*** for data transformation and cleaning.
* Implemented CI/CD pipelines using GitHub Actions to automate the testing and deployment of Python data loaders, enabling continuous integration and delivery.
* Built event-driven ETL automation with ***AWS Lambda*** function with Python that triggered the ETL process based on S3 uploads, reducing cost by 20%.
* Collaborated with data scientists and researchers to design a ML model leveraging the Neo4j database that predicts the progression of cancer based on patient’s medical history, genetic data, and other clinical information.

**Data Engineer** | United States Department of Homeland Security **May** 2021 – August 2022

* Coordinated with SME and key stakeholders to define data policy and established governance framework; developed data dictionaries, data catalogs, and data lineage documentation.
* Designed an Enterprise Data Warehouse to consolidates disparate data sources from multiple DHS agencies into a harmonized schema, providing a single source of truth for DHS data assets.
* Built interactive dashboards in Power BI leveraging the data repository, leading to more informed decision-making processes help start 2 initiatives.

**Data Analyst** | FEMA **August** 2018 – May 2021

* Constructed data pipelines with ***AWS Glue*** to seamlessly ingest and transform legacy datasets from Oracle and Postgres into ***AWS S3 data lake house***, leveraging ***PySpark*** for efficient data transformation at scale.
* Designed and implemented a complex ***AWS Athena query*** system to retrieve and process data from AWS S3 data lake, enabling faster insights for the data science team.
* Developed and deployed Spark applications on AWS EMR for large-scale data aggregation and analysis of historical disaster data, enhancing the speed and scale of data preparation for reporting.
* Collaborated with stakeholders to create standardized data models and enforce consistency across relational databases (AWS Aurora, Oracle) and non-structural repository (AWS S3, DynamoDB).
* Constructed a layered data architecture with a normalized central Layer and a staging Area; tailored data marts to meet the unique analytical needs for different reports.

**Education**

**Olivet Nazarene University** Bourbonnais, IL

Bachelor of Engineering, Major in Electrical Engineering2012 – 2016

**Technical Skill**

* **Cloud/Big Data: AWS (S3, Lambda, Glue, EMR, Athena, Aurora)**
* **Programming & Scripting: Python (PySpark, Pandas, NumPy, SQLAlchemy), SQL (T-SQL, PL/SQL), Node.js**
* **Databases: RDBMS (Postgres, Oracle, SQL Server, MySQL), NoSQL (Neo4j, MongoDB, DynamoDB), JSON, XML**
* **ETL & Data Warehousing: ETL/ELT Pipeline Design, Data Modeling, Data Governance, dbt**
* **BI & Visualization: Power BI, Tableau, Excel**
* **DevOps & Tools: Git, GitHub Actions (CI/CD), Jira, Docker**

**Certification**

* Meta Database engineer - Jan 2025
* Google Advanced Data Analytics Specialization - Mar 2024