Wesley Lau

(571) 354-9369 • wesleylau.wcl@gmail.com • [LinkedIn](#)

PROFESSIONAL EXPERIENCE

Data Engineer | National Cancer Institute | Contract   2022 – 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 | Contract   2021 – 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 | Contract   2018 – 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 distributed 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.

Data Analyst | Flagstar Bank| Contract   2017-2018
- Developed and deployed an efficient custom ETL framework that parsed complex XML structures along with JSON datasets into MongoDB; streamlined data processing workflows, saving approximately 30 hours per month.

Data Analyst | First Guarantee Mortgage Company | Contract   2016 - 2018
- Designed a comprehensive data visualization dashboard leveraging Qlik Sense to convey complicated data sets to the business Analysts for improving customer's experience.

EDUCATION

OLIVET NAZARENE UNIVERSITY   Bourbonnais, IL
Bachelor of Engineering, Major in Electrical Engineering   2012 – 2016

OTHER

Technical Skills: Data Warehouse, ETL/ ELT, Change Management, Data Governance
Tools and software: Python (PySpark, dbt, Pandas, SQLAlchemy), PL/SQL, T-SQL, RDBMS (SQL Server, Postgres, Oracle, MySQL), Non-structural data store (MongoDB, Neo4j, XML, Json), Node JS, Tableau, Power BI, Excel, AWS, Git, Docker, Jira
Certification: Meta Database engineer (Jan 2025), Google Advanced Data Analytics Specialization (Mar 2024)