Wesley Lau

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

**Professional Experience**

**Data Engineer** | National Cancer Institute | Contract2022 – 2025

* Led the integration of two external systems into the existing data architecture, expanding the Neo4j Graph data model while ensuring compatibility with downstream reporting systems.
* Developed and executed a large-scale data migration strategy (3TB), extracting data from diverse sources including Oracle and Postgres, utilizing Python ETL scripts for transformation, and implementing rigorous data validation checks to ensure data quality and integrity during loading into the Neo4j database.
* Implemented CI/CD pipelines (GitHub Actions) for automated testing and deployment of data loaders.
* Built event-driven ETL automation using AWS Lambda, reducing operational costs by 20%.
* Collaborated with cross-functional teams (data scientists, researchers) to translate research needs into technical specifications for ML model data requirements.
* Documented data pipeline architecture, ETL processes, and data models to ensure clarity and maintainability.

**Data Engineer** | United States Department of Homeland Security | Contract2021 – 2022

* Established a robust data governance framework coordinating with SMEs and stakeholders; developed comprehensive data dictionaries, catalogs, and lineage documentation to ensure data consistency and compliance across disparate systems.
* Designed and implemented an Oracle-based Enterprise Data Warehouse, consolidating diverse data sources into a harmonized schema to establish a single source of truth for DHS data assets.
* Developed reports and interactive dashboards (Power BI) providing key insights to stakeholders, supporting data-driven decision-making and the launch of new initiatives.
* Identified and resolved complex data quality issues through systematic troubleshooting and analysis.

**Data Analyst** | FEMA | Contract2018 – 2021

* Designed, developed, and maintained scalable AWS Glue data pipelines using PySpark to ingest, transform, and load diverse datasets from legacy systems (Oracle, Postgres) into an AWS S3 data lake house.
* Developed and optimized complex PL/SQL scripts and stored procedures within the Oracle database environment to support data transformation and reporting needs.
* Designed and implemented standardized data models for relational databases including Oracle and AWS Aurora, ensuring adherence to data modeling best practices and conventions.
* Performed Oracle query tuning and implemented complex AWS Athena queries, significantly improving data retrieval performance for the data science team.
* Automated and monitored data workflows (AWS CloudWatch) to ensure reliability.
* Created detailed technical documentation for data models, ETL processes, and complex query logic, facilitating knowledge sharing.
* Provided technical guidance to team members on data modeling and SQL optimization techniques.

**Data Analyst** | Flagstar Bank| Contract2017-2018

* 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 | Contract2016 - 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 Engineering2012 – 2016

**Technical Skills**

* Databases:
  + Oracle (Database Design, PL/SQL, Stored Procedures, Views, Query Tuning, Data Modeling, EDW Implementation)
  + RDBMS (SQL Server, Postgres, MySQL)
  + NoSQL (MongoDB), Graph (Neo4j)
* Data Engineering & Warehousing:
  + ETL/ELT Development, Data Integration, Data Migration
  + Data Warehousing design, Data Lakes, Data Lake Houses
  + Python (PySpark, Pandas, NumPy, SQLAlchemy), T-SQL, Node JS
* Cloud Platforms: AWS (S3, Glue, Lambda, Athena, Aurora, CloudWatch), Cloud Database Concepts (experience handling databases in AWS cloud environment)
* Data Governance & Management: Data Governance Frameworks, Data Quality Assurance, Data Modeling Standards, Data Catalogs, Data Lineage, Technical Documentation, Data Processing Flowcharting
* BI & Reporting: Power BI, Tableau
* DevOps & Tools: Git, Docker, GitHub Actions, Jira, Excel