MASTER COPY OF THINGS (NOT FOR USES!!) ☹

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

* Architected a unified Neo4j graph data model to integrate clinical trial, genomic, patient medical, and pharmaceutical research data, creating a consolidated research hub.
* 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, cleaning, and implementing data validation checks to ensure data quality and integrity.
* 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 an AWS Lambda function (Python) triggered by S3 uploads, reducing operational costs by 20%.
* Collaborated with data scientists and researchers to understand data needs and design an ML model leveraging the Neo4j database to predict cancer progression based on patient history, genetic data, and clinical information.
* Documented data pipeline architecture and ETL processes to ensure clarity and maintainability.

**Data Engineer** | United States Department of Homeland Security | Contract2021 – 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.
* Department of Homeland Security – Contract
* Collaborated with SME and stakeholders to gather data requirements and create data governance policy, established governance framework.
* Assisted in designing and developing an Enterprise data warehouse solution for housing data from different agencies.
* Developed and implemented a structured migration plan to import data from multiple data sources, including non-structural and structural data, to the centralize data repository.
* Built interactive dashboards in Tableau to monitor and validate data quality during and post-migration, providing real-time insights into accuracy, completeness, and consistency metrics.
* Engineered and executed a complex SQL query to aggregate data from centralize data repository.
* Established a robust data governance framework by coordinating with SMEs and stakeholders to define data policy; developed comprehensive data dictionaries, catalogs, and lineage documentation to ensure data consistency and compliance.
* Designed an Enterprise Data Warehouse consolidating 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 to deliver reporting results and insights to stakeholders, leading to more informed decision-making and the launch of two new initiatives.

DHS FEMA – Contract

* Designed and constructed scalable AWS Glue data pipelines to ingest, transform, and load diverse legacy datasets (Oracle, Postgres) into an AWS S3 data lake house, optimizing data processing workflows using PySpark.
* Designed and implemented complex AWS Athena queries, performing query tuning to improve performance and enable faster insights for the data science team.
* Automated and monitored data workflows using AWS CloudWatch to ensure reliability and minimize manual intervention.
* Collaborated with stakeholders to create standardized data models and enforce consistency across relational databases (AWS Aurora, Oracle) and non-structural repositories (AWS S3, DynamoDB).
* Constructed a layered data architecture with a normalized central layer and a staging area; tailored data marts to meet unique analytical needs for different reports.
* Created detailed documentation for data models and complex query processes, facilitating knowledge sharing and ensuring consistency for analytical teams.

**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.
* Engineered a robust Talend ETL pipeline to seamlessly transfer large datasets from an Oracle database into MongoDB; increased data accessibility for business analysts by ensuring integration was completed within two hours daily.

**Education**

**Olivet Nazarene University** Bourbonnais, IL

Bachelor of Engineering, Major in Electrical Engineering2012 – 2016

**Other**

**Technical Skills**:

Data Warehouse,

ETL/ ELT,

Change Management,

Data Governance ,

Serverless experience (Lambda, step functions),

Multi-Dimensional OLAP,

Star/Snowflake Schema Design

**Tools and software**:

Python (PySpark, dbt, Pandas, PySpark, 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)