

Wesley Lau

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

PROFESSIONAL SUMMARY

Accomplished Data Modeler and Engineer with over 6 years of experience designing and implementing robust data solutions for large-scale federal systems. Expertise in full-lifecycle data modeling (conceptual, logical, physical) and architecting scalable data warehouses using a variety of relational and NoSQL databases in cloud environments. Proven ability to collaborate with cross-functional teams to translate business requirements into efficient data solutions, establish strong data governance, and support modern DevOps practices.

PROFESSIONAL EXPERIENCE

Data Engineer | National Cancer Institute

NOVEMBER 2022 - January 2025

- Architected a unified Neo4j graph data model to integrate disparate cancer research systems, enhancing query capacity by 50% and providing researchers with a consolidated data source.
- 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.
- Built event-driven ETL automation with *AWS Lambda* function with Python that triggered the ETL process based on S3 uploads, reducing cost by 20%.
- Improved on a predictive cancer progression model by engineering optimized Python ETL pipelines to ingest and process terabytes of complex genomic data (FASTA/FASTQ), increasing accuracy of the model's training dataset.

Data Engineer | United States Department of Homeland Security

MAY 2021 - August 2022

- Led the design of conceptual and logical data models as the foundation for a new data governance framework, collaborating with SMEs to reduce data inconsistencies by 40%.
- Translated business requirements into a scalable physical data model for a multi-agency Enterprise Data Warehouse, enabling complex analytics by consolidating data from 5 distinct sources, including Oracle, PostgreSQL and SQL Server.
- Engineered and automated data ingestion pipelines using Informatica PowerCenter, creating reusable workflows to extract and transform data into the central data warehouse.

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.
- Accelerated data science initiatives by implementing a high-performance AWS Athena query engine, reducing data retrieval times by 60%.
- Engineered a real-time data pipeline using *Kafka* and Spark on AWS EMR to stream high-volume disaster data into Hadoop (HDFS), reducing critical report processing time by over 30%.
- Drove data consistency across the organization by developing and documenting comprehensive conceptual, logical, and physical data models for both relational (AWS Aurora, Oracle) and NoSQL (DynamoDB) platforms.
- Designed and implemented a layered data architecture (staging, normalized, and data mart layers) to meet unique analytical requirements for various stakeholder reports.

Data Analyst | 4C Decision

OCTOBER 2016 - August 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.

EDUCATION

Bachelor of Engineering, Major in Electrical Engineering

OLIVET NAZARENE UNIVERSITY | Bourbonnais, IL

TECHNICAL SKILLS

- **Data Modeling & Design:** Conceptual, Logical, & Physical Data Modeling, Data Warehousing, Data Governance, ERWin, ER/Studio
- **Databases:**
 - **NoSQL:** MongoDB, Neo4j, Elasticsearch, DynamoDB
 - **Relational (RDBMS):** PostgreSQL, MySQL, Oracle, AWS Aurora, SQL Server
- **Data Processing & ETL:** Python (Pandas, NumPy), PySpark, AWS Glue, Apache Kafka, Spark, Informatica PowerCenter, Hadoop (HDFS)
- **Cloud Platforms:** AWS (S3, Lambda, EMR, Athena, Aurora, Glue, Step Functions)
- **DevOps & Tools:** Git, CI/CD, Docker, Kubernetes

CERTIFICATION

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