Data Analyst/ Data Engineer

WESLEY LAU

(571) 354-9369 | wesleylau.wcl@gmail.com
LinkedIn Profile
Fairfax, VA

Education

BACHELOR OF SCIENCE IN ENGINEERING – Drexel University – Bourbonnais, IL   Jun 2016
Majors: Electrical Engineering

Certification

GOOGLE ADVANCED DATA ANALYTICS SPECIALIZATION   Mar 2024
Credential ID: 2FA6R2B257GJ

QLIK SENSE DATA ARCHITECT-V3   Sept 2017
Credential ID: QSDA0368v3

Technical Proficiencies

Python	Database Management	Data Visualization
PyTorch, pandas, dbt, SQLAlchemy, PySpark, boto3	Oracle, Postgres, SQL Server, MySQL	Power BI, Tableau, Excel
AWS	**Data Modeling**	**NoSQL**
S3, Athena, Glue, DynamoDB	Erwin Data Modeler Oracle SQL Developer Data Modeler	MongoDB, Neo4j, JSON, XML

Work Experience

DATA ENGINEER – National Cancer Institute – Bethesda, MD   Dec 2022 - Jan 2024
- Expanded existing Neo4j graph database to integrate incoming Postgres data from legacy system.
- Validated data post-migration through comprehensive integrity checks and created scripts to resolve discrepancies, achieving 100% data accuracy.
- Developed a Node.js API that bridged interactions between front-end application and MongoDB.
- Optimized the Python data loader system by implementing multithreading and data caching techniques, achieving a 40% increase in overall data loading speed.

DATA ANALYST – Department of Homeland Security– Arlington, VA   May 2021 - May 2022
- Collaborated with SME and stakeholders to define data requirements and translated them into actionable items, which directly contributed toward developing more robust logical and physical models.
- Engineered and executed a complex SQL query to aggregate data from centralize data repository, reducing report generation time by 60%.
- Monitored critical data quality metrics and enforced validation rules, established a thorough governance standard that led to a reduction in migration errors by 30%, ensuring high-quality datasets.
- 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.

DATA ANALYST – DHS FEMA – Washington D.C.   Aug 2018 – May 2021
- Developed data pipelines using PySpark to ingest and transform legacy data from Oracle and Postgres and load them into centralize AWS S3 data lake.
- 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.
- Collaborated with stakeholders to create standardized data models and enforce consistency across relational databases (AWS Aurora, Oracle) and NoSQL stores (AWS S3, DynamoDB).
- Conducted in-depth data profiling and established rigorous quality checks, guaranteeing reliability for analytics workflows utilized by the data science team; enhanced accuracy metrics to 99% across machine learning models.

DATA ANALYST – Flagstar Bank – Detroit, MI   Oct 2017 – Mar 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.
- Created comprehensive ER diagrams to support business requirements and improve communication between stakeholders and the technical team.
- Provided infrastructure for user authentication and access token within Qlik.
- Designed interactive data visualizations for 230 non-technical employees, simplifying complex data sets and facilitating better decision-making, increased data comprehension and actionable insights across the organization.

DATA ANALYST – First Guarantee Mortgage Company – McLean, VA   Sept 2016 – Sept 2017

- Leveraged SQL, Python, and Qlik Sense to analyze 20TB of data and improved business decisions making.
- 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.