

Albert Wei

(203) 529-7086 | New York, NY | alwei@umich.edu | weialbert.github.io

Education

University of Michigan

MSE Industrial & Operations Engineering – Operations Research

Ann Arbor, MI

BSE Industrial & Operations Engineering

Experience

Palantir Technologies

Data Engineer

2025—present

New York, NY

- Centralized data quality monitoring, dataset scheduling, and health check alerting fleetwide across all Foundry products and Apollo deployments
- Implemented Iceberg table generation and procedures into Foundry platform, enabling self-service dataset generation and product monitoring capabilities across 100+ product teams
- Migrated petabyte-scale production datasets to Iceberg table format, reducing storage and query costs by 40%+ and enabling real-time read capabilities for Foundry product usage datasets (ie. cloud cost, token usage, and deployment performance)
- Generated executive-level product usage performance metrics used directly by CFO in quarterly earnings reports and investor communications
- Designed and maintained ingestion and load balancing pipelines for edge deployments with irregular internet connectivity, ensuring reliable data delivery and end-to-end visibility

Capital One

Senior Data Analyst

2021—2025

McLean, VA

- Refactored legacy retail bank overdraft risk model pipeline, consolidating 8 cloud data sources and cutting runtime from 4+ hours to 20 minutes via parallelization and Apache Spark optimization
- Owned US Card complaints data pipeline (Snowflake, Python), migrated data source to S3 enabling real-time updates, reduced query time from 300s to 1s, and led change management for 100+ analysts onto production system serving SVP+ leadership
- Designed real-time recoveries analytics infrastructure using Airflow, Apache Hamilton, and Snowflake supporting \$10M+ in core business decision-making
- Led federal Community Reinvestment Act (CRA) compliance data strategy, coordinating cross-functional teams and delivering regulatory submissions to the US Treasury OCC
- Built central data quality framework in Python and Apache Airflow, replacing 20+ adhoc solutions and scaling to 70+ production datasets with advanced statistical trend analysis

University of Michigan – Transportation Research Institute

Model Developer & First Author

2018—2021

Ann Arbor, MI

- Automated landmark and feature extraction from 1,000+ CT images with Machine Learning radial basis functions and cubic spline interpolation
 - Published peer-reviewed first-author paper on parametric head geometry modeling using PCA and multivariate linear regression in MATLAB
-

Technical Skills

- **Programming & Core Tools:** C++, Rust, Python (pandas, numpy, scikit-learn, TensorFlow, PyTorch), SQL (Snowflake, PostgreSQL, NoSQL), R, CI & CD (Github Actions, Jenkins)
- **Data Engineering & Infrastructure:** Apache Spark, Airflow, Kafka, Hamilton, AWS (S3, Glue, Redshift, EMR, SageMaker, Lambda, Athena), Azure (Databricks, Blob Storage, Data Factory, Synapse), GCP (BigQuery, Cloud Storage, Dataflow, Dataproc), Docker
- **Analytics & Visualization:** Jupyter Notebooks, Tableau, Power BI, Statistical Regression Analysis, Time Series (ARIMA, Prophet), A/B Testing
- **Machine Learning & Optimization:** Predictive Modeling, Natural Language Processing, Anomaly Detection & Variance Testing, Linear & Nonlinear Optimization, Operations Research