

Jacky Lam

Senior Data Engineer

ny.jackylam@gmail.com | 646-287-5229 | linkedin.com/in/jackylam-io | jackylam.io

PROFESSIONAL SUMMARY

Senior Data Engineer who builds data platforms that scale gracefully with the business. I design orchestration, CI/CD, reverse ETL governance, and custom connector architecture as one connected system — unlocking meaningful gains in compute efficiency, deployment confidence, and developer velocity. An analytics background keeps me close to how data is actually used, so every platform decision stays anchored to the outcomes it's meant to drive.

TECHNICAL SKILLS

- **Languages:** Python, SQL
- **Cloud Platforms:** AWS, GCP
- **Data Warehousing & Databases:** Snowflake, Redshift, BigQuery, PostgreSQL
- **ETL/Orchestration:** dbt, Apache Airflow, Fivetran, Airbyte
- **Big Data Tools:** Databricks, Apache Spark
- **DevOps/CI-CD:** Git, Github Actions
- **BI/Visualization:** Looker, Tableau

EXPERIENCE

Education First

Boston, MA

Senior Data Engineer

August 2025 - Present

- **Event-Driven Orchestration Redesign:** Audited and re-architected the Airflow orchestration layer from fragmented cron-based scheduling to a fully event-driven pipeline with Fivetran as the single data freshness trigger. Replaced 30+ cron DAGs with Dataset-chained sensors across 18 Fivetran connectors, unified dbt, Hightouch, and Iterable into a single coordinated pipeline, and eliminated race conditions, wasted compute, and stale data propagation.
- **Talon.One Connector Architecture & Reliability:** Designed a custom Fivetran connector using the Connector SDK to close an analytics gap with no native integration. Architected a declarative, registry-driven framework with per-scope cursor isolation, partial-failure containment, and rate-limit-aware retries. Delivered multi-environment CI/CD via GitHub Actions with AWS OIDC for safe promotion across dev, QA, and production.
- **dbt CI/CD Pipeline Optimization & Branch Isolation:** Diagnosed five compounding inefficiencies in the dbt CI/CD pipeline spanning redundant compute, shared dev schemas causing developer collisions, and unnecessary database dependencies in static analysis. Designed branch-based schema isolation in Snowflake with automated cleanup on merge, consolidated duplicate CI jobs into a single dbt build pass, and decoupled lint workflows from Snowflake connectivity to cut lint times by ~85%.
- **Reverse ETL Environment Isolation & CI/CD:** Identified critical production risk in the reverse ETL platform where marketing and data teams built and tested directly in production. Designed and implemented a dev-first workflow with dedicated Hightouch environments, Git-backed version control via YAML sync, and PR-gated promotion to production, eliminating the risk of accidental campaign sends and establishing an auditable change management process.

- **AI-Augmented Developer Workflow Automation:** Automated the commit, code review, and pull request lifecycle across repositories using AI-augmented workflow tooling. Engineered structured commit generation, severity-classified code review, end-to-end PR creation with JIRA integration, and parallelized batch review of open PRs, reducing manual overhead in the release cycle while enforcing consistent engineering conventions.

Nuts.com

Jersey City, NJ

Data Engineer

June 2021 - August 2025

- **ETL Pipeline Refactor & Cost Optimization:** Refactored and optimized the core ETL data pipeline ingesting all website sales data. Rewrote transformations and re-architected batch jobs to reduce infrastructure costs by 63% and cut job run times by 50%.
- **Data Platform Modernization & dbt Architecture:** Migrated notebooks and scripts into Git-based version control, implemented CI/CD pipelines for Databricks, and architected dbt-based data modeling with fact and dimension tables, standardizing deployments and reducing query times.
- **Stabilized BI Infrastructure with Blue/Green Deployment:** Engineered and deployed a Blue/Green data deployment strategy to stabilize BI infrastructure, reducing query latency errors by over 80% and significantly improving dashboard reliability and user trust.
- **Product Cost Metrics:** Drove a company-wide initiative to unlock visibility into product cost metrics by leading cross-functional collaboration between Finance, BI, and DevOps. Spearheaded project scoping, requirements gathering, and the end-to-end development of scalable data pipelines in Databricks and reporting infrastructure in Looker, enabling more accurate, actionable financial insights.

Motivate LLC.

New York, NY

Senior Data Analyst

July 2019 - June 2021

- **Post-Acquisition Data Platform Migration:** Architected and executed a data warehouse migration following Lyft's acquisition, reconciling legacy systems with Lyft's modern infrastructure and leading a parallel BI platform consolidation from Domo to Tableau across 8 North American cities.
- **Technical Team Leadership & Cross-Functional Delivery:** Led a cross-functional data team through concurrent platform migrations, providing technical direction on warehouse architecture, data continuity strategy, and reporting infrastructure modernization.

NYC Taxi & Limousine Commission

New York, NY

Senior Policy Analyst

March 2016 - July 2019

- **Defined Regulatory Data Standards and Vendor Compliance:** Authored regulatory data specifications and led vendor compliance efforts. Partnered with IT to implement data ingestion pipelines, ensuring accurate and timely loading of for-hire vehicle datasets into production systems.
- **Launched High-Impact GIS Tool Viewed Over 2.3M Times:** Redesigned and digitized the Department of Transportation's For-Hire Vehicle Relief Stand database using GIS tools, producing an interactive, driver-friendly map viewed over 2.3 million times since launch, dramatically improving public accessibility.

EDUCATION

Queens College, City University of New York

Queens, NY

Bachelor of Business Administration

2014

MILITARY SERVICE

New York Army National Guard

New York, NY

Air Battle Management System Operator

Mar 2013 - Mar 2016