

Junjie (Caspar) Chen

Fremont, CA | (207) 286-4242 | jc5067@columbia.edu | [caspar0-0.github.io](https://github.com/caspar0-0)

EXPERIENCE

Taskrabbit, Inc.

San Francisco, CA

Marketplace Platform — IKEA Group · terabyte-scale Snowflake data warehouse

Data Engineer II

Nov 2022 – Present

- **Architect** and maintain a multi-layer Snowflake warehouse (staging → intermediate → mart) using dbt, with enforced contracts, column-level lineage, and unit-tested semantic models serving Finance, Marketing, Operations, and executive reporting.
- **Drive a dbt Semantic Layer initiative** consolidating fragmented job metric YMLs into a unified fact/dimension model (Star Schema), including ISO-week Last Year (LY) comparison logic and multi-market company metrics — enabling self-serve metric querying and reducing definition drift across teams.
- **Led data infrastructure buildout** for an acquired company (Dolly), designing a greenfield dbt repository from scratch — staging, intermediate, and mart layers — to integrate acquired data assets into Taskrabbit's unified warehouse.
- **Engineered modular SQL transformations** via CTEs, Jinja macros, and dbt incremental strategies, cutting warehouse compute costs by 30%+ through query profiling, partition pruning, and clustering optimization.
- **Built and operate Airflow (Astronomer) DAGs** with dynamic task mapping, cross-DAG dependencies, and fail-safe retry logic for daily production pipelines processing millions of marketplace events.
- **Integrated multi-source data** from REST APIs (Stripe, Fixer, Segment, Attom), S3, and internal microservices via Fivetran and custom Python connectors — enforcing schema contracts across 160+ production tables and delivering end-to-end data products via dbt models on top of third-party real estate data.
- **Resolved a \$4.4M billing variance** (down to \$14K) through deep-dive IKEA billing discrepancy analysis across four European markets (UK, ES, PT, FR) — identifying systemic tracking errors and formula mismatches across accounting workflows.
- **Automated Stripe vs. CSI revenue reconciliation** by mapping granular transaction flows (charges, transfers, refunds) into Snowflake, building Python scripts to automatically extract and flag transaction anomalies, integrating platform earnings categories to enable end-to-end financial integrity checks.
- **Contributed to Snowflake IaC** using Terraform, implementing RBAC design for data warehouse privilege management across engineering and analytics teams.
- **Architected and deployed company-wide Claude AI integrations** — connecting Claude to Snowflake (service role + prod token), dbt Cloud, Sigma, and Atlassian via MCP — and built a fully automated daily metrics pipeline computing KPIs (AOV, Take Rate) with YoY/WoW comparisons and anomaly detection delivered to Slack stakeholders.
- **Implemented data observability** via unit tests, Snowflake dynamic data masking, column tagging, and quality alerting — reducing mean incident response time by 40%.

Data Engineer I

Sep 2021 – Oct 2022

- **Led migration** of legacy data models to a modern stack (Fivetran, dbt, Snowflake, Looker) — reducing ETL pipeline runtime from 8 hours to under 30 minutes while maintaining data quality.
- **Triaged and resolved** urgent production ETL issues, maintaining SLA compliance and data availability for downstream analytics and reporting consumers.
- **Developed dbt SQL models and Airflow DAGs** powering marketing attribution pipelines and operations reporting — enabling campaign performance tracking and operational KPI visibility across business teams.

Apple (via Apex Systems)

Sunnyvale, CA

Data Analyst Lead

Dec 2020 – Sep 2021

- **Performed ETL** from multiple geospatial sources to modernize Apple Maps — building ingestion pipelines using QGIS, SQL, and Scala into the data warehouse.
- **Validated and patched** territory data with PostgreSQL and Hive on Hadoop; delivered data ingestion for 8 countries from OpenStreetMap with quality and efficiency.

EDUCATION

Columbia University

New York, NY

M.A. in Statistics

Sep. 2018 - Dec. 2019

Coursework: Data Mining, Databases, Data Science, Time Series Analysis, Nonparametric Methods

University of Wisconsin–Madison

Sep. 2014 - May. 2018

B.S. Applied Mathematics, Actuarial Science

SKILLS & CERTIFICATIONS

Data Engineering: dbt · Snowflake · Airflow (Astronomer) · Fivetran · Terraform · Python · SQL · Scala

Cloud & Infrastructure: AWS · S3 · Segment · Mixpanel

BI & Analytics: Sigma · Looker · Tableau

Developer Tools: GitHub · Cursor · Claude AI (MCP) · Jira · Confluence

Certifications: AWS Certified Cloud Practitioner, Astronomer Certified DAG Authoring for Apache Airflow, DataCamp: Cleaning Data in Python, Web Scraping in Python