

# MINH PHAM

Email | [LinkedIn](#) | [Portfolio](#) | [GitHub](#) | 682-597-4780

## Experience

---

**Insurify** · Insurance aggregator · \$600M Series B · \$50M→\$800M ARR hypergrowth May 2025 – Present  
Analytics Engineer – Marketing & Growth Partnerships Boston, MA

- Owned **GTM metrics infrastructure** across **30 paid channels** in a **1200-model dbt warehouse**, enabling decisions across Marketing, Growth, Finance; shipped **15 Slack alerts** and **metric-consistency tests**, resolved **200 tickets** & merged **300 PRs**
- Designed **attribution** and **spend pipelines** for **Google + Microsoft SEM** via **Hightouch reverse ETL, GAQL, Airbyte, Redshift, dbt**; refactored the search-performance dbt layer with **shared macros**, lifting margin **20%**
- Built a **TV multi-touch attribution model** (**\$50M+/yr**) tagging branded vs non-branded SEO revenue via a **Google Search Console branded-rate dbt model** (**Airbyte → S3, Redshift**) feeding the MMM for **media-mix** optimization
- Built **paid acquisition data product** for **The Trade Desk** (**\$50M+/yr** across CTV, audio, display), ingesting TTD APIs into **S3, Redshift** via **Airbyte + Terraform**, modeling **dbt marts** in **5 self-serve reports** with **per-brand parallel streams**
- Investigated **cross-channel spend discrepancies** across **13 partner networks** using **end-to-end data lineage**; reconciled **LTV/CAC across channels** in **Mode**, recovering **\$500K** and improving **unit economics** by **25%**
- Replaced manual partner onboarding with an **Everflow pipeline** streaming clicks and conversions through **Kinesis** into a **Redshift materialized view**, dbt-parsed with **Airflow WAP**, scaling to **150+ partners** and **\$10M+ revenue**
- Refactored a **300GB+ RTB bidding dbt model: non-SARGable filters** → **anti-joins**, pruned upstream deps, tuned **DISTKEY** and **SORTKEY** – cutting **storage 100GB** and **runtime 50%**
- Built **Partnerships Spend Bot**, an **LLM-automated workflow** pulling **Asana** tickets and running **Claude** sessions to generate dbt seed PRs and post to **Slack**, replacing **3 engineers'** weekly cycles; productionizing on **Claude API + dbt MCP**

**Lazard** · Elite boutique investment bank & asset manager · NYSE: LAZ · \$240B+ AUM May 2024 – August 2024  
Data Engineer Intern – Financial Advisory Business Technology New York, NY

- Migrated a **Monte Carlo simulation** from Jupyter to a **fault-tolerant PySpark pipeline** on **AWS EMR**, orchestrated via **Airflow** with **retryable, idempotent tasks**, reducing runtime by **40%** and productionizing daily revenue forecasts
- Built **5 Streamlit data products** backed by **Alteryx APIs** and **Snowflake**, improving self-service analytics for **30 bankers**
- Designed a **Snowflake analytics model** in **dbt** with standardized transformations, tests, and docs for BI consistency

**TechSmith** · Visual communication software (Snagit, Camtasia) · 30M+ users January 2024 – May 2024  
Business Intelligence Engineer Intern – Marketing & Sales East Lansing, MI

- Engineered a **real-time, event-driven pricing pipeline** using **Kafka** to ingest Salesforce data into **Databricks**, where **PySpark MLlib** and **TensorFlow** processed **5M+ records daily**, improving accuracy by **18%**
- Deployed **dbt models** in **Snowflake** to unify fragmented datasets, enabling a **single source of truth** for **20+ dashboards**
- Built scalable **ETL pipelines** with **PySpark** and optimized **SQL** to support **GTM reporting** for Marketing and Sales, applying **incremental processing** to cut ad-hoc turnaround time by **50%**

## Projects

---

**Skytrax Reviews Extract-Load Pipeline** | *Airflow, S3, Snowflake, Terraform, Python* | [GitHub](#)  

- Scraped **160K+ airline reviews** via **26 parallel A-Z tasks** with **dynamic task mapping** into versioned, encrypted **S3**
- Chained **3 DAGs** via **Airflow Datasets** with idempotent **Snowflake COPY INTO**; **S3, IAM, Snowflake** all via **Terraform**

**Skytrax dbt Warehouse & CI/CD** | *dbt, Snowflake, Airflow, GitHub Actions, AWS, Terraform* | [GitHub](#)  

- Built a **Kimball star schema** (5 dims, 1 fact); **slim CI + defer + favor-state CD** for incremental deploys
- Keyless OIDC** from **GitHub Actions** to **AWS**; full **Terraform IaC** for **Snowflake RBAC**; **dbt docs** on **CloudFront**

**FinPulse-Fraud — Lambda Architecture** | *Spark, Kafka, Flink, Pinot, PrestoDB, Airflow, Docker* | [GitHub](#)  

- Built a **Lambda architecture: Spark batch** over **HDFS** plus **Flink stream scoring** with **exactly-once 2PC** on **Kafka**
- Dual-serving: Pinot hybrid table** (sub-second dashboards) + **PrestoDB-on-HMS** (granular ad-hoc SQL); **Superset BI**

## Technical Skills

---

**Data Warehouse/Lakehouse:** Snowflake, Redshift, BigQuery, Databricks, DuckDB/MotherDuck, Iceberg, Delta Lake, Trino, PrestoDB, Pinot, Hive, HDFS

**Data Processing & Transformation:** Apache Spark, Spark Structured Streaming, Flink, Alteryx, Kafka, Kinesis, CDC, dbt Core

**Data Integration & Reverse ETL:** Airbyte, Fivetran, dltHub, Hightouch

**Orchestration:** Apache Airflow, Astronomer, Prefect, Mage, Dagster, Orchestra

**BI & Analytics:** Mode, Hex, Power BI, Tableau, Looker Studio, Streamlit, Visivo, Lightdash, Superset, Datadog

**DevOps & Cloud:** Terraform, Docker, Kubernetes, Git, GitHub Actions, AWS, GCP

**AI Tooling:** Claude Code, Cursor, Codex, MCP servers (dbt, Asana, Slack)

## Education

---

**Texas Christian University — GPA: 3.74 (Magna Cum Laude)** Aug 2021 – May 2025  
Computer Systems Analysis & Minor in Mathematics, Fintech Fort Worth, TX

- Awards:** Program Technology Award, Dean's List