

YUN-CHI PANG

Boston, MA | 669-274-6138 | yunchipang@gmail.com | linkedin.com/in/yunchipang | github.com/yunchipang

PROFESSIONAL EXPERIENCE

Software Engineer II, Data Platform @ Klaviyo 01/2026–
Global SaaS marketing automation platform powering personalized email and SMS for 183K+ brands

- Built company-wide Spark and Iceberg tables on AWS EMR for all internal analytics and ML pipelines
- Designed Kafka and Flink streaming workflows to maximize S3 resource utilization and reduce storage costs
- Engineered fault-tolerant Airflow (MWAA) ETL pipelines, optimizing data partitioning and compute efficiency
- Led cross-team initiatives with data infra peers to deliver unified, production-grade internal datasets

Data Engineer @ Dashmote 01/2022–06/2022
Series B SaaS data analytics company serving Fortune 500 clients across 50+ global markets

- Deployed automated AWS data pipelines processing 10M+ records daily, reducing total system runtime by 88%
- Architected a scalable Python and Scrapy framework to ingest 1M+ global menu records per day
- Resolved critical data inconsistencies by collaborating across engineering and QA, cutting error rates by 30%

TECHNICAL SKILLS

Data Infra & Storage: Kafka, Flink, Spark, Iceberg, Hudi, HDFS, ClickHouse, Hive, Trino

Cloud & Platform: AWS (EMR, S3, EC2), GCP, Docker, Kubernetes, Terraform

Languages & Backend: Python, Java, Rust, SQL, Scala, FastAPI, Spring Boot, Git

OPEN-SOURCE

Apache Kafka  15+ PRs 03/2025–

- Eliminated tech debt across 200K+ deployments by migrating core Scala modules to Java, reducing contributor onboarding time by 50% and simplifying long-term maintenance
- Slashed CPU overhead by 20% and memory allocations by 60% in config API through algorithmic optimization and OOP refactoring—leveraged polymorphic overloading and stream-based lazy evaluation

Apache Hudi  10+ PRs 07/2025–

- Accelerated selective queries by 10× via column statistics data skipping, file-level pruning reduces scans by 90%
- Achieved 100× speedup for time-range queries through automatic timestamp-to-partition filter transformation
- Parallelized incremental processing (8× faster) via file slice partitioning APIs for distributed ETL pipelines

Apache Airflow  10+ PRs 03/2025–

- Reduced enterprise attack surface by shifting CLI from DB access to API-driven authentication

Apache Gravitino  20+ PRs 04/2025–

- Extended REST APIs (Java/Python/CLI) enabling cross-system queries across HDFS, Hive, and Iceberg
- Built unified authorization layer enabling centralized RBAC management across data sources

Additional contributions across major data-infra ecosystems (ClickHouse, OpenLineage, Sedona, OpenDAL, Ruff), focusing on reliability, performance, and developer experience.

EDUCATION

Northeastern University, MSc in Computer Science 09/2022–05/2025

The Chinese University of Hong Kong, BBA in Business Analytics & Quantitative Marketing 09/2017–12/2021