

YUN-CHI PANG

Santa Clara, CA | 669-274-6138 | yunchipang@gmail.com | linkedin.com/in/yunchipang | github.com/yunchipang

EDUCATION

Northeastern University, MSc in Computer Science	2022–2025
The Chinese University of Hong Kong, BBA in Business Analytics	2017–2021

TECHNICAL SKILLS

Languages & Backend Python, Java, C++, Rust, SQL, Django, FastAPI, Spring Boot, Node.js


Data Technologies: Kafka, Spark, Flink, Airflow, Iceberg, Hudi, ETL/ELT

Databases: MySQL, PostgreSQL, MongoDB


Cloud & DevOps: AWS (EC2, S3), GCP, Docker, Kubernetes, Jenkins, GitHub Actions

Developer Tools: Git, Pytest, JUnit, Postman, Swagger/OpenAPI

OPEN-SOURCE

Apache Kafka  [10+ PRs](#) *Event streaming platform used by 80%+ of Fortune 100* 2025–


- Migrated legacy **Scala** to **Java**, modernizing core modules for 1250+ contributors and 200K+ deployments
- Improved response construction latency by 10% and reduced config-related CPU usage by 15% by eliminating redundant collections and repeated method calls

Apache Gravitino  [20+ PRs](#) *Federated metadata lake used by Uber, Pinterest, Xiaomi, and more* 2025–

- Unified metadata authorization with permission pushdown for **relational DBs**, **data warehouses** and **lake catalogs**, cutting API-layer auth logic by 60%
- Built extensible **REST APIs** for **Java**, **Python**, and **CLI** clients, enabling federation across **Hadoop**-compatible systems (**HDFS**, **Hive**, **Iceberg**) and boosting query performance for 45K+ monthly users

Apache Airflow  [10+ PRs](#) *Workflow orchestration platform used by 77K+ orgs* 2025–

- Reduced enterprise attack surface by shifting CLI from DB access to API-driven authentication
- Unblocked release of new Airflow provider versions by resolving critical mypy type-checking issues

OpenLineage  *Open standard for data lineage adopted by Databricks, Astronomer, and more* 2025–

- Ensured **Spark 4** backward compatibility via reflection-based factory patterns to handle constructor changes across 20+ connectors including Kusto and Databricks

*Additional contributions to emerging and community-driven projects, including [Apache Hudi](#), [Apache Fluss](#), [Apache Uniffle](#), [ClickHouse](#), [Ruff](#), and [Greenstand](#), focusing on enhancements, bug fixes, and improving developer experience.

PROFESSIONAL EXPERIENCE

Software Engineer Intern – OurDate 2023

AI-powered growth-stage social platform focused on event-driven matchmaking

- Built scalable **REST APIs** (**Django**, **MySQL**) powering core matchmaking features, improving system throughput by 25% and reducing latency for 100+ daily active users
- Led peer code reviews, mentoring junior engineers and raising code quality, cutting post-release bugs by 20%

Data Engineer – Dashmote 2022

Series B SaaS data analytics company serving Fortune 500 clients across 50+ global markets

- Built and scaled automated **AWS** data pipelines processing 10M+ records daily, reducing runtime by 88%
- Developed a distributed scraping framework (**Python**, **Scrapy**) to ingest 1M+ restaurant menu records per day, expanding product coverage by 3x across global markets
- Resolved critical data inconsistencies by collaborating across engineering and QA, cutting error rates by 30%

Ops Intern – ShopBack 2020–2021

Regional e-commerce scale-up with 35M+ users across Asia-Pacific

- Automated workflows by replacing manual tasks with **Python** tools, increasing ops efficiency by 90%
- Eliminated 98% of support tickets by implementing Selenium-based tracking, driving 1% weekly growth