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   09/2022–05/2025
The Chinese University of Hong Kong, BBA in Business Analytics   09/2017–12/2021

TECHNICAL SKILLS

Languages & Backend: Python, Java, Rust, SQL, Django, FastAPI, Spring Boot, Node.js
Databases & Storage: MySQL, PostgreSQL, MongoDB, Hive, Iceberg, HDFS, ClickHouse
Cloud & DevOps: AWS (EC2, S3), GCP, Docker, Kubernetes, Jenkins, GitHub Actions
Testing & Developer Tools: Git, Pytest, JUnit, Postman, Swagger/OpenAPI

OPEN-SOURCE

Apache Kafka 🌐 *15+ PRs* *Event streaming platform used by 80%+ of Fortune 100*   03/2025–
- Modernized core modules by migrating **Scala** to **Java**, improving maintainability across 200K+ deployments
- Removed redundant data structures and calls, improving response time by 10% and cutting config-related CPU use by 15% through algorithm and **memory optimizations**
- Enhanced log compaction validation by adding compression parameters and extending test coverage by 30%

Apache Airflow 🌐 *10+ PRs* *Workflow orchestration platform used by 77K+ orgs*   03/2025–
- Reduced enterprise attack surface by shifting CLI from DB access to **API-driven authentication**

Apache Gravitino 🌐 *20+ PRs* *Federated metadata lake used by Uber, Pinterest, Xiaomi*   04/2025–
- 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
- Unified metadata authorization with permission pushdown for **relational DBs**, **data warehouses** and **lake catalogs**, cutting API-layer auth logic by 60%

Apache Hudi 🌐 *5+ PRs* *Data lake platform used by Uber, Amazon, ByteDance*   07/2025–
- Boosted scalability by adding APIs for file slice splitting and incremental change processing, enabling parallel reads across 1000+ partitions and cutting query latency by 40% on large datasets

OpenLineage 🌐 *Open standard for data lineage adopted by Databricks, Astronomer*   07/2025–
- Ensured **Spark 4** backward compatibility via reflection-based factory patterns to handle constructor changes across 20+ connectors including Kusto and Databricks

PROFESSIONAL EXPERIENCE

Software Engineer Intern – OurDate   07/2023–10/2023
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   01/2022–06/2022
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

Ops Intern – ShopBack   09/2020–07/2021
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