

# YUN-CHI PANG

Santa Clara, CA | 669-274-6138 | [yunchipang@gmail.com](mailto:yunchipang@gmail.com) | [linkedin.com/in/yunchipang](https://linkedin.com/in/yunchipang) | [github.com/yunchipang](https://github.com/yunchipang)

## 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 |

## TECHNICAL SKILLS

Languages & Backend: Python, Java, Rust, SQL, Django, FastAPI, Spring Boot, Node.js  
Distributed Systems & Data Infra: Kafka, Airflow, Spark, Flink, Iceberg, HDFS, ClickHouse  
Databases: MySQL, PostgreSQL, MongoDB  
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* 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* *Data lake platform* 07/2025–
- Introduced column-stats data skipping, reducing scanned data by 90% and accelerating selective queries by 10×
  - Reduced query latency by up to 100× via automatic timestamp-to-partition filter transformation, eliminating 95% of unnecessary scans for time-range queries
  - Added file-slice partitioning APIs enabling 8× faster incremental processing across distributed ETL workers
- Apache Airflow** 🔄 *10+ PRs* *Workflow orchestration platform* 03/2025–
- Reduced enterprise attack surface by shifting CLI from DB access to API-driven authentication
- Apache Gravitino** 🔄 *20+ PRs* *Federated metadata lake* 04/2025–
- Developed REST APIs (Java, Python, CLI) to enable cross-system queries across HDFS, Hive, and Iceberg, improving performance for 45K+ monthly users
  - Unified metadata authorization with permission pushdown across databases, warehouses, and lake catalogs, reducing API auth logic by 60%

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

## PROFESSIONAL EXPERIENCE

- Software Engineer Intern @ OurDate** 07/2023–10/2023  
*AI-powered growth-stage social platform focused on date-planning*
- Boosted system throughput by 25% by developing scalable REST APIs (Django, MySQL) powering core date-planning features—reduced latency for 100+ daily active users
  - Suppressed post-release defects by 20% through systematic peer code reviews and mentoring junior engineers
- Data Engineer @ Dashmote** 01/2022–06/2022  
*Series B SaaS data analytics company serving Fortune 500 clients across 50+ global markets*
- Built and optimized data collection AWS pipelines processing 10M+ records/day, cutting runtime by 88%
  - Designed scalable scraping framework (Python, Scrapy) ingesting 1M+ menu records/day across global markets
  - Resolved critical data inconsistencies by collaborating across engineering and QA, cutting error rates by 30%
- Ops Intern @ ShopBack** 09/2020–07/2021  
*Regional e-commerce scale-up with 35M+ users across Asia-Pacific*
- Increased ops efficiency by 90% by automating manual workflows with Python tools
  - Cut 98% of support tickets through Selenium-based tracking automation—drove 1% weekly partner growth