

CHAN Yao Ying

☎ +65 97609802 | ✉ me@chanyaoying.com | 🔗 [chanyaoying](https://chanyaoying.com) | 🌐 chanyaoying.com

Professional Experience

Data Engineer - GovTech Singapore, Data Programme [Straits Times Feature] June 2023 - Present

- An engineer working on the national data analytics platform, enTRUST, bringing real world research and academic impact. It is a high-throughput, air-gapped and highly secure analytics platform.
- Achieved 81.48% runtime reduction and 97.69% cost savings by designing and optimizing Spark-based ETL workloads
- Built and productionized a secured big-data delivery system leveraging on an event driven architecture, DynamoDB (single-table design), and serverless compute, enabling reliable transfer of datasets up to hundreds of GB
- Led end-to-end development (from requirements gathering) of a metadata catalogue with a self-hosted open source search engine, improving data discoverability and accessibility to fit-for-purpose datasets; designed for horizontal scalability
- Reduced build and deployment time by 50% by redesigning GitLab CI/CD pipelines with caching strategies and parallel jobs, accelerating release cycles
- Rapidly reduced tech debt by modernizing codebases for AI-assisted development, using context engineering and agentic workflows to automate large-scale refactoring and documentation standardization across multiple repositories
- Tuned and debugged production applications across distributed environments, reducing failure rates and improving operational SLAs
- Mentored interns to help them deliver production-ready features
- Drove interoperability of different products within the department, coordinating with product managers and engineers across teams
- Drove system hardening efforts that resulted in zero critical vulnerabilities during penetration testing

Data Engineering Intern - GovTech Singapore, Data Engineering Jan 2023 - April 2023

- Developed DIAB (Data Infrastructure as Code) scripts and documentation using Terraform to accelerate agencies' adoption of cloud-based data platforms and data warehouses on the Government Commercial Cloud
- Added new AWS service modules (e.g., S3, Glue, EMR) and configurations for data-centric use cases; remediated security findings to pass penetration testing
- Enabled teams in at least four agencies to reduce infrastructure deployment timelines from 6 months to 1 month by providing robust templates and hands-on support

Education

Singapore Management University Singapore

BSc | Information Systems (Business Analytics) & Computing Studies (Artificial Intelligence) ~ August 2019 - May 2023

Skills

Certifications:

- AWS Certified Solutions Architect - Associate
- AWS Certified Cloud Practitioner
- VMware Certified Technical Associate - Data Centre Virtualization
- Certified Scrum Master® (CSM®)
- HashiCorp Certified: Terraform Associate (003)
- ISTQB® Certified Tester Foundation Level (CTFL)

Programming Languages: 🐍 Python (proficient), 📘 TypeScript (proficient), 🟡 Go (learning)

Tools and Frameworks: React, Next.js, OpenAPI, Terraform, Spark, Kafka, Hadoop, SQL, NoSQL (DynamoDB, MongoDB, Redis), ETL/ELT pipelines, AWS Well-Architected Framework, GitLab CI/CD, Docker, Kubernetes, Jenkins, Git, basic ML with tensorflow, pytorch, pandas and PySpark

Languages: English (native), Chinese (proficient)

Personal Projects

Expense Management Telegram Bot (2025)

- For users in Telegram groups to manage and simplify debts (like in Splitwise)
- Integrated with custom PayNow QR for quick debt settlement
- **100% AI-generated**, leveraging on GitHub's spec-kit
- Active usage in 3 groups (pilot)
- AWS Lambda, DynamoDB, S3 Static Hosting - to keep costs minimal

Realtime Analysis of Fashion Trends via Tweets (2023)

- Built an end-to-end pipeline for real-time text analysis of Twitter streams using Apache Kafka, Spark Streaming and MongoDB, with interactive dashboards in Tableau via Apache Drill
- Leveraged Tweepy API to ingest social media data; implemented sentiment and trend analysis to visualize changing fashion conversations

Algorithmic Pairs Trading (2023)

- Low-frequency statistical arbitrage trading system that identifies cointegrated stock pairs, trains models validated by backtesting and continuously retrains with new data.
- Built with Flask, Redis, MongoDB

Working Overseas

Eligible for the H-1B1-Singapore Visa for working in the United States