Jason Ho

ccjasonho@gmail.com | https://www.linkedin.com/in/jason-ho-b77194255/

Work Experience

Qube Research & Technologies Apr 2026 - Oct 2026
Software Engineering Internship, **Commodity Data** Team London

- Built the **data infrastructure**, including **distributed compute** and **storage**, to support **idempotent**, **fault-tolerant** pipelines for geospatial data, handling both **live and historical backfill** workloads at **terabyte scale**
- Designed an **extensible ETL framework** that ingests **multi-dimensional** data from **heterogeneous** legacy-format sources into a centralised **data lake**, applying access-pattern-specific optimisations to deliver **significantly faster query performance** than prior methods
- Developed **downstream pipelines** that automate **data transformations** on incoming data, delivering reliable, ready-to-use datasets for researchers to access

Man Group Jun 2025 - Sep 2025
Software Engineering Internship, **Machine Learning Technology** Team London

- Designed and built an **MCP Registry and Client**, extending internal AI systems for **MCP compatibility** and enabling **1500+ employees** to discover, configure, and integrate **enterprise tools** (e.g., Jira, Confluence) directly into **AI assistants**
- Developed full-stack features including **CRUD APIs, authentication, usage tracking, and pinning** for improved usability
- Developed and optimised **reinforcement learning pipelines (GRPO, LoRA)** for large language model fine-tuning in trading-signal generation, achieving an **80% improvement in Sharpe ratio** through tailored reward function design

Spendsy Limited Jul 2024 – Sep 2024
Full Stack Developer Internship Hong Kong

- **Led a cross-functional team of 5** to build a data-driven dashboard using **PowerBI**, **Azure**, and **Terraform**
- Focused on **access management** by obtaining the **service principal's access token** and **report's embed token**, and implementing **row-level security** (RLS) in the semantic models to restrict data access based on user roles
- Designed and built a **mobile coupon attachment verification system**, allowing customers to upload attachments to S3 and retrieve them via **S3 URLs** for business owner verification, leveraging **Nodejs, AWS Lambda**, and **DynamoDB**

Education

Imperial College London, MEng Computing (Artificial Intelligence and Machine Learning) Sep 2023 – Jun 2027

- **Dean's List** (First-Class Honours)

Projects (https://github.com/Jho0413)

GraphDB, In-memory graph database engine in **Java**

- Engineered **concurrent ACID transactions** with **MVCC-based Snapshot Isolation**, providing non-blocking reads over consistent graph snapshots and **optimistic** write-conflict detection across parallel writers
- Implemented **crash-safe durability** via checksummed **Write-Ahead Logging (WAL)**, recovering all committed transactions after simulated crashes
- Built a modular graph query engine with algorithms including **Dijkstra, Bellman-Ford, Tarjan, Johnson, and Floyd-Warshall**, accelerated by **LRU caching**

FlashLoom, AI flashcard generator website [**View**]

- Created a **full-stack web** that generates flashcards autonomously from text, YouTube links, and PDF uploads, utilising **Gemini LLM** with **Pinecone** for embeddings and **Retrieval-Augmented Generation**
- Engineered **scalable, production** infrastructure using **NextJS, Firebase, Inngest (Background processing), Stripe (subscriptions), and Clerk (authentication)**, enabling secure payments, seamless onboarding, and efficient background task processing for flashcard creation

PintOS, Operating System in **C**

- Collaborated in a **team of 4** to implement kernel-level features including **priority scheduling with donation**, **synchronisation primitives**, **user program execution** with system calls and process management, and **file descriptor–based I/O handling**
- Built **virtual memory support** with frame allocation, eviction, swapping, lazy loading of executables, dynamic stack growth, and memory-mapped files

Skills

Languages: Python, Java, C, JavaScript/TypeScript
Data & Cloud: AWS (ECS, Lambda, S3), Dask, Docker, Pandas, Polars, PostgreSQL, Xarray, Zarr
Tools: CI/CD, Git, Jupyter, Linux, Postman, pytest