Lim Jing Jie

✉ jj.limmm159@gmail.com 📞 +65 8121 4356 in linkedin.com/in/lim-jing-jie ⌸ jjlim7 🔗 jjlim7.com

EDUCATION

Bachelor of Science in Software Engineering   Aug 2022 – Jun 2026 (expected)
Singapore Management University

WORK EXPERIENCE

Software Engineer Intern   May 2025 – Jul 2025
Millennium Management ⌐
- Developed a Rule Management System using Java Spring Boot, enabling portfolio managers to define and execute custom data validation logic prior to report generation, reducing data quality issues by 30-50%.
- Engineered a modular rule evaluation framework leveraging object-oriented design principles and patterns (Strategy, Registry), enabling scalable rule registration and execution with minimal developer overhead
- Integrated with Spark-based pipelines to support post-report validation across large datasets, improving operational efficiency and reducing manual intervention by the support team

Big Data Engineer Intern   Jan 2025 – May 2025
Binance ⌐
- Built scalable, low-latency data pipelines using Kafka, Flink, and Spark to power real-time search indexing and enrich token metadata with CoinGecko data, expanding coverage of searchable tokens by 40%, significantly improving content discoverability on Binance Square.
- Designed and implemented a Spark-driven data quality monitoring system leveraging heuristic rules to detect anomalies and trigger automated alerts, reducing false positives by half, significantly improving monitoring precision.

Software Engineer Intern   Nov 2024 – Jan 2025
Tokka Labs ⌐
- Engineered and optimized an onchain monitoring system in Rust, leveraging asynchronous programming and concurrency to increase transaction tracking accuracy by over 25% across multiple business lines, including RFQ and CEX-DEX arbitrage.
- Implemented a high-performance, event-driven backend that achieved sub-100ms event processing latency and reliably handled thousands of trade events per second to support secure and scalable DeFi operations.

**Data Engineer**   Aug 2023 – Sep 2024
Secretlab
- Facilitated dimensional modeling and optimized ELT processes with dbt, Fivetran, and Snowflake, achieving a ~18% decrease in data processing time and enhancing data accuracy by 5%.
- Collaborated across departments to build a feature store and deploy a machine learning model for freight delivery date estimation, resulting in a 12% reduction in delivery delays.
- Deployed an ETL pipeline to ingest data from partner APIs using AWS Lambda, increasing reporting speed by 10%.

Software Engineer   Jun 2021 – Apr 2023
Scantist ⌐
- Designed and deployed a scalable data processing pipeline utilizing Google Cloud Dataproc, Apache Spark, Airflow, and Python, to automate ingestion of CVE and package manager data (NPM, PyPI), delivering critical vulnerability intelligence powering the SCA platform with 200% throughput improvement.
- Automated ETL and web scraping workflows to process over 100 million records from security advisories and open-source registries, saving up to 8 hours of manual work daily and ensuring a timely and accurate vulnerability data.
- Implemented rigorous testing frameworks using pytest with 80%+ test coverage across all data pipelines, significantly enhancing data quality and accuracy.

PROJECTS

UniExplorers [github link] ⌐
- Developed a centralised university exchange platform empowering students to seamlessly explore exchange programs and connect with peers through an interactive forum.
- Architected backend microservices in Go and NodeJS, employing Kubernetes for container orchestration.
- Built a real-time gRPC notification service with Go and RabbitMQ for enhanced user engagement and responsiveness.

PROFESSIONAL CREDENTIALS

- AWS Certified Solutions Architect Associate ⌐   • GCP Certified Professional Data Engineer ⌐

SKILLS

Coding: Go, Rust, TypeScript, JavaScript, Python, Java (Spring Boot), SQL │ **DevOps:** Git, Docker, Kubernetes, Terraform, GitHub Actions │ **Data:** Airflow, Spark, Hive, MongoDB, Elasticsearch, BigQuery, Redis, Snowflake, dbt