Jonathan Wang

DATA ENGINEER · SOFTWARE ENGINEER

✉ jonnowangnz@gmail.com | jonnowang | jonathan-wang-nz

Experience

Marshall Wace *London, United Kingdom*
DATA ENGINEER *Apr. 2024 - Present*
- Onboarded over **30 new ETL pipelines across quant, fundamental, operations and compliance teams.** Datasets included TB scale alternative data such as weather forecasting as well as traditional equity, derivative and fixed income datasets. (LSEG, S&P, Markit etc...)
- **Optimised Excel to database interface, reducing on-load latency from 2 minutes to 5 seconds.** Employed multiple strategies including: keep-warm caching, request batch scheduler, javascript sorting optimisation and proxy upgrades leveraging HTTP/2 concurrency.
- **Reduced the latency of the central authentication API by 100x from 5s to 50ms P95 response time.** Built an API layer that reconstructed and cached the Active Directory graph with a sharded Redis cluster to reduce recursive search space and remove redundant calls.
- Built an index rebalance insight parser leveraging generative LLMs. Articles were converted from unstructured HTML to JSON with correctness and schema validation stages. **Delivered structured insights to quant teams databases within seconds of article publication.**

Smarkets *London, United Kingdom*
DATA ENGINEER & SCIENTIST *Sep. 2022 - Apr. 2024*
- Developed an industry-leading recommendation engine. Processed user activity data in an implicit matrix factorisation model; hyperparameter tuning was done using a custom metric suite. A/B testing revealed a **40% increase in engagement** compared to unpersonalised events.
- Implemented an asynchronous Python job using Kafka which loaded archived S3 files without severe I/O bottlenecking for trading backtesting. New system **reduced processing time for terabytes of CSV data to 12 minutes** per day of data, down from 2 hours previously.
- Worked closely with internal DevOps team to migrate Rust and Nix pricing engine data systems from outdated infrastructure to a unified architecture. **Reduced the engineering maintenance effort by 50%** due to unified code architecture.

HomeX *Cambridge, United Kingdom*
APPLIED DATA SCIENCE INTERN *Jun. 2021 - Sep. 2021*
- Built a backend API using Python's FastAPI serving a regression model to 5 other internal services.
- Built a front-end BI data visualisation dashboard using TypeScript and React with interactive street maps and tables.
- Deployed these services using Docker, Kubernetes and Terraform with data ingestion from AWS S3 and Snowflake.

Education

University of Cambridge *Cambridge, United Kingdom*
M.ENG (HONS), COMPUTER AND INFORMATION ENGINEERING *Oct. 2018 - Jun. 2022*
- Graduated **Honours with Distinction** (1st Class thesis & 1st Class exam results).
- First Year: **74%** ; Second Year: Ungraded (As a result of CoVid-19) ; Third Year: **71%** ; Fourth Year: **72.5%**.
- **Masters Thesis**: Developed a large language model that could respond to hate speech using a novel evaluation suite and custom data. Modified cutting edge models such as Blender Bot & GPT-3 to achieve superior performance. Training was done on HPC clusters.

Westlake Boys' High School *Auckland, New Zealand*
5A* AT A LEVEL *Feb. 2013 - Dec. 2017*

Skills

Programming Python, Typescript, Linux, Git
Data SQL, Large Language Models, Pandas/Polars, Airflow, AWS S3, Snowflake, SQL Server, PostgreSQL, Redis, Kafka
DevOps AWS, GCP, Docker, Kubernetes, CI/CD, LGTM Monitoring

Projects

Conference for Truth and Trust Online *Boston, Massachusetts*
FEATURED SPEAKER *Oct. 2022*
- Gave an on-stage presentation to a 100+ person audience regarding the novel technical contributions to LLM research based on my Master's thesis. (https://truthandtrustonline.com/)

Honors & Awards

2021	**College Prize for Engineering**, Robinson College	*Cambridge, UK*
2019&20	**College Exhibition Awards for Academic Achievement**, Robinson College	*Cambridge, UK*
2017	**New Zealand Outstanding Scholars - Top 50 in Country**, NZQA	*Auckland, NZ*