

SKILLS

- Data Engineering
- AWS
- Databricks
- dbt
- Python
- SQL
- Terraform
- CI/CD

Certifications

- 12/2021 AWS Certified Solutions Architect – Associate
- 07/2023 Databricks Certified Data Engineer – Professional

WORK EXPERIENCE

Cascade Strategy, Sydney

Data Engineer

01/2022 – Current

- Currently serving as the sole data engineer on the team, responsible for end-to-end data engineering tasks, managing everything from data ingestion to infrastructure, modelling, integrations, and incident management.
- Played a key role in building the AWS-based data platform from scratch and developing the entire data stack for the company.
- Developed self-hosted data ingestion/integration tools to migrate from third-party services, saving the data team 50% in operational costs.
- Designed and implemented real-time and batch data ingestion services from a variety of sources (e.g., SNS, Kinesis, S3, Webhooks, API etc.) into the data platform
- Solely built, maintained, and optimized the data integrations from the data platform to the BI, CRM, billing, and notification systems via reverse-ETL.
- Applied extensive data transformations and data modelling (i.e., developing medallion data models using dbt) to enhance data accuracy and insights
- Worked collaboratively with business teams to develop business intelligence models and dashboards.
- Implemented principles of CI/CD and Infrastructure as Code for efficient delivery and maintenance of the data infrastructure and services.
- Ensured engineering best practices were followed: version-controlled all codes, enforced testing for jobs and data models, and utilized performance-optimized data approaches.
- Crafted and maintained useful utilities for the data team, including IDE (VSCode) extensions, dbt macros, and multi-system integration tools.
- Swiftly resolved operational incidents, minimizing downtime and maintaining data integrity

The University of Melbourne, AU; The University of North Carolina at Chapel Hill, USA

Postdoctoral Research Fellow

09/2017 – 12/2021

- Undertook research projects in computational biomechanics, human-machine interaction(wearable robotics), and applied machine learning (reinforcement learning, computer vision)
- Supervised PhD/master students and delivered lectures/tutorial sessions.
- Helped non-technical researchers in their data or computational challenges
- Published over 10 research papers in peer-reviewed journals. These can be found on my Google Scholar profile at, <http://surl.li/iybmj>

PROJECT

Data Infrastructure, Ingestion, Modelling, and Integration System

2021 - Current

- This comprehensive system provisions data infrastructure and orchestrates services for data ingestion, modelling, and integration.
- As the sole caretaker of this system, I maintain the data infrastructure and am responsible for the design and implementation of data ingestion and integration services.
- I have contributed to over 75% of the data modelling/transformation repository, as measured by both commits and line counts in the dbt repository.
- The system is highly scalable and has been in production for over a year with nearly zero downtime.
- I host data ingestion and integration services in-house, which has allowed us to replace costly third-party services such as Fivetran and Census, resulting in a significant reduction of data operational costs by over 50%.
- View the system architecture here, <http://surl.li/iybuh>

Multi-region Sensitive Backend Data Stream/Batch Ingestion

2023 - Current

- This project involves a detailed implementation of the backend data ingestion pipeline.
- It sanitizes sensitive data within the source regions due to data security regulations, and globally ingests them into a single region.
- This solution replaces the existing single-region data ingestion process that lacked data sanitizing.
- The pipeline supports both real-time streaming and batch ingestion for data backfilling.
- The implementation fully incorporates principles of Continuous Integration/Continuous Deployment (CI/CD) and Infrastructure as Code (IaC).
- View the service architecture here, <http://surl.li/iybvb>

dbt Databricks Power User – A VSCode extension

07/2022

- An open-source extension that enhances the integration of VSCode with dbt and data platforms (e.g., Databricks).
- This serves as a comprehensive replacement for dbt Cloud Development in a local environment, providing additional features.
- The extension has been publicly downloaded over 2000 times since its publication.
- Check it out on the VSCode marketplace, <http://surl.li/iybni>

Run & Attack – An Online Multiplayer Mobile Game

2021–2023

- An online player-versus-player mobile game, developed using the Unity platform, integrates GPS-tracked running activity with a board game.
- This game aims to promote a healthier lifestyle through engaging gameplay.
- The backend was developed using AWS services, including DynamoDB, Lambda, Cognito, and GameLift, and was deployed using CI/CD and Terraform.
- Innovative use of Amazon GameLift resulted in a more than tenfold reduction in operational costs.
- Check out the demo on YouTube, https://youtu.be/QCSvZ_SCai8

University Research Projects Series

2017 - 2021

- Successfully conducted a series of university research projects in the fields of computational biomechanics, applied machine learning, computer vision, and human-machine interactions.
- These projects have resulted in numerous successful research funding and paper publications.
- For more details, please visit my personal homepage at <https://melbww.com/>

EDUCATION

The University of Melbourne

Ph.D. of Engineering

2013–2017

- Melbourne International Research Scholarship

- Melbourne International Fee Remission (100% fee waiver)

Tianjin University

Bachelor of Engineering - 3.80 GPA

2008–2012

- Caterpillar Inc. Scholarship,
- Samsung Co., Ltd. Scholarship
- National Scholarship

Nankai University

Bachelor of Economics (Finance), dual degree

2009–2012