

Ted Zhao

10/4 Sheppards Lane, Norwood SA 5067, Australia | ted.zhao.au@gmail.com | 0416 551 622

linkedin.com/in/ted-zhao | github.com/tade0726

Summary

- Senior Data Engineer with proven expertise in optimising data architectures, reducing schema size by 50% and query runtime by 50%.
- Extensive experience with modern data stack (DBT, Dagster, Databricks) and cloud platforms (AWS, Azure).
- Skilled in real-time data streaming and batch processing, implementing production systems with Redis, Elasticsearch, and PySpark.
- Track record of building scalable data solutions from simulation systems to RAG-powered chatbots.

Experience

Data/AI Engineer, Rainmakr.ai – Adelaide, SA (Remote) May 2023 – July 2024

- Simplified 30+ table data models with DBT, reducing schema size by 50%.
- Reduced redundant data storage by 15% through deduplication and date range alignment.
- Optimised queries through re-indexing and join refinements, cutting runtime by 50%.
- Implemented SCD principles and incremental updates, saving 10% in ingestion time.
- Integrated DBT with Dagster Cloud for mixed orchestration of notebook and DBT jobs.

Machine Learning Engineer, Skyscanner – Shenzhen, China Nov 2019 – July 2021

- Developed T+1 batch processing pipeline using AWS Lambda/PySpark for feature extraction, storing in Parquet format.
- Engineered real-time feature streaming from Elasticsearch with Redis caching, improving inference CTR by 2%.
- Designed Databricks-based data lake architecture for model training and real-time inference.
- Implemented Redis-based caching layer for rapid data access and real-time feature serving.

Data Scientist, DJI – Shenzhen, China Apr 2018 – Aug 2018

- Built Scrapy-based web crawler to gather city demographics and competitor store distribution data for market analysis.
- Developed quantitative ranking algorithm incorporating economic indicators and business factors for city prioritisation.
- Analysed post-launch user demographics (location, income, age) to create detailed customer segments.
- Monitored sales patterns and provided data-driven insights for campaign strategy optimisation.
- Developed Django/SQLAlchemy-based API for Jupyter extension, enabling efficient SQL querying and data model management.

Data Scientist, Smart Decision Technology Ltd – Shenzhen, China Aug 2015 – Mar 2018

- Implemented credit scorecard system with feature selection, WOE calculation, and logistic model conversion in Python.
- Built SimPy-based logistics simulation with Flask/SQLAlchemy backend, securing 2-year contract with S.F. Express.
- Applied K-Means clustering for HKJC member segmentation, delivering lifecycle-based targeting strategies.
- Developed Chinese text sentiment analysis using Word2Vec and K-means for thematic clustering.
- Built supervised learning model with custom Chinese tokenization for app review classification.

Projects

ATO Chatbot 2023

- **Live Demo:** ato-chat.streamlit.app | **Source:** github.com/tade0726/ato_chatbot

- Built production-grade RAG system with LlamaIndex and Qdrant vector store, enabling context-aware tax information retrieval.
- Implemented MLOps pipeline with ZenML for data ingestion, OpenAI fine-tuning, and automated model evaluation.
- Developed Streamlit-based UI with real-time query rephrasing and source citations for enhanced user trust.

Quantitative Trading with Reinforcement Learning – University of Adelaide Sept 2022

- Researched and developed crypto trading strategy using Reinforcement Learning at University of Adelaide.
- Designed multi-asset portfolio optimisation with Sharpe Ratio reward system for buy/sell quantity decisions.
- Built dual Deep Learning modules for price signal prediction and market sentiment analysis.
- Implemented comprehensive backtesting framework evaluating Sharpe Ratio and Maximum Drawdown (MDD).
- Developed profit analysis system incorporating trading fees and compound returns visualisation.

Languages

English: IELTS 7.0 **Mandarin:** Native **Cantonese:** Native

Education

University of Adelaide, Master in Machine Learning – Adelaide, SA June 2021 – May 2023

Shenzhen University, Bachelor in Mathematical Sciences – Shenzhen, China Sept 2010 – July 2014

Certifications

Coursera, Machine Learning - Oct 2016
<https://www.coursera.org/account/accomplishments/verify/Q89DMB9RYYTB>

Udacity, Machine Learning Nanodegree - Dec 2018
<https://graduation.udacity.com/confirm/PKAQPLU2>

Skills

Data Engineering: DBT, Databricks, Snowflake, Dagster

ETL: SQL, SQLAlchemy, Spark, Pandas

AI/Machine Learning: Reinforcement Learning, Deep Learning, Clustering, LightGBM, XGBoost, SVM, Word2Vec, A/B Testing

MLOps/Cloud: Azure ML, AWS, SageMaker, MLflow

Programming Languages: Python, VBA, Scala, Julia

Web Development: Django, Flask, Asyncio, Scrappy

Version Control & Containers: Git, Docker