

Kuan Pin (Ben), Chen

+1 415 918 8446; kchen91@dons.usfca.edu; <https://www.linkedin.com/in/kuan-pin-chen/> <https://github.com/benntuecon>

EDUCATION

University of San Francisco,

MS, Data Science

San Francisco, CA

Jun 2023(*Expected)

- Courses: Data Structures/Algorithms, Distributed Data Systems, Relational Database, NoSQL, Advanced Machine Learning, MLOps, Data Engineering*, A/B Testing (experiment design).

National Taiwan University,

BA, Economics

Taipei, Taiwan

Jun 2019

SKILLS

Programming Languages: Python, Javascript, SQL, Java, Scala, Go(Golang).

Infrastructure/FallStack: Kubernetes(k8s), Docker/Docker-compose/Docker Swarm, Git, GitHub Action, Jenkins, AWS, GCP, RabbitMQ, gRPC, RPC, Vim, Linux, Shell, Jenkins, Django, Fastapi, React.js, Node.js, Nginx.

Data Engineer/Machine Learning: Databricks, Airflow, Kafka, Spark/PySpark Spark Streaming, PyTorch (PyTorch Lightning), Scikit-learn.

Certifications: AWS Machine Learning Specialty (MLS-C01), AWS Certified Solutions Architect – Associate (SAA-C02)

WORK EXPERIENCE

USF Data Institute (with Amazon Machine Learning Solutions Lab)

Machine Learning Engineer Intern

San Francisco, CA

Oct 2022- June 2023

- Designed and researched optimization algorithms for accelerating large-scale deep learning training, mentored by Amazon Machine Learning Solution Lab (AWS MLSL) and USF Data Institute.
- Focused on the sampling strategy including selective back-propagation and importance sampling, and the attention-layer-emphasized sampling strategy on transformer models.

Chun Hong Web.

Software Engineer

Taichung, Taiwan

May 2021- Mar 2022

- Built a customer relationship management system with Django for clients to monitor and analyze consumer activities.
- Developed an interactive chatbot with grammy.js, Line API, and Auto.js, automizing customer services.
- Designed relational database schemas with Entity-Relationship Diagram (ERD), implement python REST APIs, and built a PostgreSQL database to achieve complex join with Django Object Relational Mapping (ORM) model.
- Designed a simple micro-service architecture system to partially scale up the worker service, increasing system scalability and robustness, and implemented a task queue manager via Bull.js, a caching service via Redis, a task consumer service with Puppeteer and Selenium, and a dashboard service with python flask and Dash(Plotly).

Chuan Jia Bao Ltd.

Backend Engineer, Co-founder

Taipei, Taiwan

Apr 2019- Apr 2021

- Built social media with Express.js, MySQL, AWS Beanstalk, AWS Elasticache, and Docker, reaching daily-active-user of 1000+ and monthly-active-user of 9000+. Selected by one of the giant accelerators in Taiwan, AppWorks.
- Modeled an NLP recommendation system via Bert as a Service, prolonged average user sessions by 40+%, and reduced content skipped rate by ~7%.
- Optimized SQL query by redesigning the query and indexation, decreasing query time by 60%.
- Initiated automation project for web scraping with selenium, and Node.js, saving manpower by 90%.

SELECTED PROJECTS

Portfolio website <https://www.benchen1996.info> (React.js, Nginx Proxy Manager).

Sep 2022

Micro-service practice, Personal Project (Golang, RPC, gRPC, and RabbitMQ)

Mar 2023

- Extended portfolio website functionality with micro-service architecture, including Frondend, Logger, Mailing, Listener, Message broker, Load-balancer, Mongodb, ML worker for visitor activity analysis, and fronting with Nginx. Services are communicating via a combination of AMQP, gRPC, RPC, and REST API coordinated by Kubernetes(k8s).

Apple Watch data ETL project, USF Project (PySpark, MongoDB, Airflow, Databrick, GCP-GCS)

Mar 2023

- Managed the data fetching process with Apache Airflow, warehoused in Google GCS, transformed then stored in the document-base MongoDB, and utilized Databricks Spark cluster to perform distributed model training and analysis.

Apache Spark Data Engineering on IMDB dataset, Personal project (Scala, Spark)

Mar 2023

- Applied functional programming in Scala to write distribute-friendly code and utilized Apache Spark as cluster manager, and MLlib to perform ML pipeline.