Roy

roy@gmail.com | +86 152 1673 6907. C,C++,Go,Java,Python,JavaScript

Professional Experience

Senior Server Engineer, nVidia

December 2019 – Present, Shanghai

Executed and built a high-performance, high-reliable advertising engine handling over 300k peak QPS.

- Reduced server latency below 90ms by rewriting the ads search engine in C++ using the bRPC framework.
- Reduced ad retrieval time by 50% by redesigning and optimizing indexing algorithms using DNF and RoaringBitmap.
- Decreased the memory size of DNN models by 20% by compressing embedding float features to half applied both in the ad engine and KV server.
- Boosted model predictions by 40% by AVX, SSE vectorization, parallel computing, and hash optimization.
- Established running analytics and metrics visualization, alerts, and logging eco.
- Yielded ad platform income by exploiting ad diversity, budget pacing, ANN recall before model ranking, etc.

Technologies: C++, bRPC, perf, valgrind, HDFS, Hive, KV, Redis.

Server Engineer,Meta

July 2019 – November 2019, Shanghai

Executed and designed an advertising ecosystem for the online agriculture-focused million-user shopping startup.

- Built a high-performance ad engine for advertisers and small merchants handling millions of requests.
- Containerized multi-microservice for ad search, ad conversion tracking, and ad charging in Go by Kubernetes.
- Improved ad prediction model AUC by 20% by training an XGBoost model to replace naïve LR.

Technologies: Go, Java, Kafka, Redis, Hive, Spark.

Full-stack Engineer,Google

June 2015 – June 2019, Shanghai

Developed and maintained the ad order ecosystem for thousands of advertisers, agencies, and merchants.

- Accelerated average release frequency by decoupling the monolith architecture to backend and frontend modules by using Laravel, and Vue.js with Webpack, and migrating to RESTful APIs.
- Eliminated repeated wheels invention in the front end by pioneering and delivering an ad-specified UI component lib by using Vue.js to save team time.
- Lessened report latency by 100% by partitioning large DB, indexing, and optimizing long queries.

Technologies: PHP, Laravel, jQuery, Vue.js, MySQL, Redis, Nginx, Apache.

Education & Language

Shanghai University, School of Computer Engineering and Science. September 2012 – April 2015

Master of Computer Applications (MCA), GPA:3.59/4.00

English: Professional working proficiency