YUAN GU

guyuan002@gmail.com ⋄ 804.874.1712 ⋄ linkedin.com/in/gu-yuan ⋄ San Jose, 95134

WORK EXPERIENCE

Software Engineer, Google - Privacy Sandbox on Android   03/2022 - Now, Mountain View, CA

I worked for the On-Device Personalization (ODP) team, which offers public **Android APIs** for app developers to serve ads without cross-app identifiers. I built large-scale, distributed, and **privacy-preserving ML systems**.
Server: a scalable service on **GCP** that orchestrates Federated Learning (FL) tasks across Android devices.

- Built robust cloud **micro-services** in **Java/Springboot**: a *TaskManager* that reads and writes FL tasks in a **Spanner** database, a *TaskScheduler* that allocates tasks to idle Android devices, and an *Aggregator* that performs FedAvg on model updates. Set up replicated model storage in **GCS** with **CDN** enabled, and **CI/CD** pipelines using **Docker**. Scaled the service to handle requests from **1+ million** devices per day.
- Developed a **Python/Flask** web service from scratch, enabling adopters to create FL tasks by customizing **TensorFlow** models and learning algorithms with **differential privacy** guarantee (e.g. DP accounting). Implemented **risk mitigation** strategies, including the use of SavedModel to prevent code injection and the deployment of a confidential space to attest that source code cannot be modified to override privacy policies.

· **Client**: Android system modules in **Java/Kotlin** with data collection, ads serving, and model inference.

- Developed solutions to enforce regulatory compliance (GDPR, COPPA, LAT) by building **framework APIs** and **services** in Android to gather privacy signals (e.g. age, user consent) from various system modules. **Xfn'ly** aligned with PM/legal to finalize privacy policies. Awarded **2000$ cash bonus** by senior leadership.
- Built resource-efficient **job schedulers** in Android and **ETL** pipelines to periodically collect user data. Designed **SQLite** table schema, **data recovery** mechanisms, and a **data access layer** with **TTL controls**.

Software Engineering Intern, Cloudera - Storage Infra Team   05/2021 - 08/2021, Santa Clara, CA

- Enriched features for Hadoop's distributed file system, Apache Ozone, by adding **Java/Maven** modules and unit tests. Resolved 10+ **Jira** issues and contributed **10k+** lines of code as an open-source committer.
- Developed dashboard systems (CLI tool, Web UI, and APIs) for Ozone on the monitoring server, Recon, that helps clients visualize disk usage (DU) of Ozone clusters, which increases observability of the system.
- Designed a **RocksDB** schema to store real-time metadata updates by taking snapshots of Ozone datanode manager (OM) via **gRPC**. Implemented **Java RESTful APIs** that query metadata from RocksDB and (recursively) compute disk usage on Recon's backend, which saved 90% **CPU cycles** on OM.
- Visualized DU as pie charts on Web UI by building frontend in **React**, **TypeScript**, **Less**, and **Plotly.js**.

Teaching Assistant, Carnegie Mellon University   06/2021 - 12/2021, Pittsburgh, PA

- Redesigned the cloud storage project for cloud computing course, which built a social networking website with heterogeneous storage in **MySQL** for login info, **MongoDB** for comments, and **Neo4j** for social relations.
- Migrated cloud infrastructure from **GCP** to **Azure**. Built **Terraform** and **shell** scripts to automate VM deployment and dependency installation, which allowed students to focus on core learning objectives.

Full-Stack Mobile Developer, Iowa State University   04/2020 - 07/2020, Remote

- Developed a cross-platformed mobile app in **Flutter/Dart** for the university's COVID-19 Tracker project's dashboard, which provides daily infected cases and death tolls. Our work has been trended on news.
- In agile iterations, managed **UI design**, **automatic location pinpoint**, **OAuth integration**, and **screen size adaptation**. Loaded a **Firebase** database for the app from raw data in CSV and JSON.

EDUCATION

Carnegie Mellon University   Pittsburgh, PA
Master of Science in Information Networking   08/2020 - 12/2021
Coursework: Computer Systems, Cloud Computing, Distributed Systems, Database Systems, Storage Systems
Honors: Merit-based Tuition Scholarship, 2020 vGHC Full Scholarship, Teaching Assistant for Cloud Computing
College of William and Mary   Williamsburg, VA
Bachelor of Science in Computer Science and Mathematics, GPA: 3.87/4.00   08/2017 - 05/2020
Honors: *summa cum laude*, James Monroe Scholar, Research Assistant fellowship, Dean's List student