

YUAN GU

guyuan002@gmail.com ◇ 804.874.1712 ◇ [linkedin.com/in/gu-yuan](https://www.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](#). My work built full-stack, distributed **privacy-preserving ML systems**.

Server: a federated compute service on **GCP** that coordinates collaborative ads reporting and model training.

- Built several mission-critical **micro-services**: a task assignment service in **Python/gRPC** that schedules compute tasks to devices, a **Spanner** database to persist task metadata, an aggregation service that performs **FedAvg** in hardened OS ([TEE](#)), and a key management server for **data encryption** and **attestation**.
- Deployed a distributed store for ML models in **GCS** buckets by **sharding** the model into fine-grained objects such as model graph, checkpoint, and gradient, which makes the system more **fault-tolerant** and **trackable**.
- Leveraging **load balancing**, **database indexing/replication**, and **scheduling algorithms**, scaled the service to handle requests across **100+ million** devices/day with **2000+ QPS** and reliable database reads.

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

- Built end-to-end, resource-efficient **ETL** pipelines to periodically collect on-device, real-time user data for **3+ billion** users. Designed and implemented **SQLite** database schema, **upgrade/rollback** mechanism, and **data access layer** to store privacy-sensitive data with **TTL controls** that clean up stale records.
- Engineered a solution that enforces privacy regulations ([GDPR](#), [COPPA](#), [AADC](#), [LAT](#), etc.) by implementing **services** and **job schedulers** in Android to obtain privacy signals from [GmsCore](#). **Xfn'ly** aligned with PM/legal to finalize privacy policies. Awarded **1000\$ cash bonus** by senior leadership.

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 **Firestore** 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