# Wanyu Zhang

3650 McClintock Ave, Los Angeles, CA 90089
+1 323-491-4325 | wanyuzha@usc.edu | linkedin.com/in/wanyu-zhang-usc

## EDUCATION

**University of Southern California** — Los Angeles, CA
*Master of Science (thesis on distributed systems(fast scaling)), Computer Science, GPA: 3.91* — *Jun 2022 – Dec 2024*

**Hefei University of Technology** — Hefei, China
*Bachelor of Engineering, Computer Science (with Hons.)* — *Sep 2016 – Jul 2020*

## PROJECTS

**Flash Burst Computing** | supervisor: Seo Jin Park — Sep 2023 – Dec 2024

- Research on Flash Burst System: enable application to use a large portion of a shared cluster for short periods of time
- (SoCC'24 Submitted)**Towards Real-time Data-Intensive Computing on Public Cloud**: Measured multiple overheads of Spark job in large scale environment on Google Cloud, and investigate potential transaction coordination overheads
- Integrate the logical process from Nu(NSDI'23) to facilitate resource fungibility in query execution and enable rapid scaling and scheduling in Spark; Research on the new shuffle algorithm from Millisort to reduce shuffle latency; Leveraged insights from the Velox engine to analyze shuffle stage and made new scaling up strategies.

**Raft KV Storage** | Go — Jan 2024 – Mar 2024

- Implemented a distributed system in Golang based on the Raft paper, including key features like leader election, log replication, fault recovery, and log compaction.
- Built a distributed KV (key-value) storage engine using the Raft protocol to ensure consistency. Realized data sharding and automated failover to address fault tolerance requirements.

## WORK EXPERIENCE

**ByteDance** — Shenzhen, China
**Backend Engineer Full Time** — *May 2021 – Jun 2022*

- **Stickers Income:** Developed and launched offline rendering stickers on Capcut in Asia and Americas regions. Peak QPS of service reached 80. Acquisition reached 220K and activation reached 3.06M.
- **Backend Service Development:** Participated in development of offline cloud rendering framework, and developed two reusable components of task scheduler with C++, Golang and ffmpeg. Developed a general pipeline with standard API for different stickers. Utilized edge computing for real time rendering, encapsulated RTC callback functions into SDK, maintained worker status through gRPC heartbeat, and decreased service latency of each frame to 200ms.
- **DevOps:** Deployed and operated hundreds of instances in total with gitlab, Cmake, SCM, Jenkins and cloud platform and performed stress testing on multiple types of GPU. Maintained usage of CPU/GPU of single instance under 70%, service time of generating a video under 3000ms and video size under 3MB. Established error monitor system on twenty types of error with automatic alerts using Kafka and visualized on Grafana.

**ByteDance** — Beijing, China
**Big Data Engineer Full Time & Intern** — *Dec 2019 – May 2021*

- **Data Warehouse Development**: Constructed data pipeline of Fanqie Novel with Spark and Python. Performed ETL process on 120B tracking event data daily from Kafka and relational databases to HDFS and constructed interaction modules (including behaviours such as comment, reply, like etc.) of data warehouse on four layers.
- **DevOps:** Deployed and operated 60+ Spark tasks with ByteDance distributed system and Hive on Spark. Created multiple SLA standards for every task based on data lineage, tagged three priority levels on tasks and assigned to different YARN queues. Established standards of cold storage for resource management. Created 60+ alarms on data quality and task latency to ensure data meets SLA standard.
- **Spark Task Optimization:** Estimated elapsed time of each stage, found and solved possible data skew in shuffle process with SparkUI and DAG graph and decreased large task computation time from two hours to thirty minutes. Collected SparkUI reports of key tasks to modify memory of executors and drivers to maintain load balancing of YARN queue. Decreased alarm number per day from 30 plus to less than 10.

## SKILLS

**Languages:** C++17/20, Java, SQL, Golang, Python, Shell

**Developer Tools:** Git, Docker, Cmake, OO Design, Distributed System, Hive, Cloud Platform, VS Code