

Guosong Li

+86 13917099196 | +1 8582813274 guosongli88@gmail.com Settle, WA | Shanghai, CN

<https://github.com/KafkaCat> | <https://www.linkedin.com/in/guosong-li-9b7637123/>

SUMMARY

- I consider myself to be keenly drawn to new technologies and innovations. I fully embraced with AI and am incredibly passionate on learning new things every day.
- 5 years of backend-focused experience designing, scaling, and operating petabyte level database management systems in cloud with data compliance.
- Experience leading diversified 5-person team to finish complex software design and deliver product on time.
- Experienced practitioner using AI coding agent to boost development/testing efficiency.

EXPERIENCE

Tonbo.IO

2025.10 - 至今

Database Engineer

- Develop the next-gen arrow/parquet native LSM-tree storage engine for full serverless OLAP architecture that stores all data in object storage
- Full asynchronously high-performance architecture storage engine which can be run anywhere, async runtime like tokio or browser or edge server. Full ACID support with no-disk architecture. All data is stored in parquet in your object storage.
 - Design and implement the read/write path including working with parquet sst, predicate pushdown, prediction-based query planning and execution, data pruning, MVCC filtering, streaming etc.
 - Implement the object storage based manifesto storage that practice serialization isolation utilizing S3's CAS feature which is a prerequisite for serverless instance.
 - Participate in implementing the WAL on S3 and compaction handling (including remote compaction) of the storage engine.
 - Help delivered the 0.4.0 public version of tonbo <https://github.com/tonbo-io/tonbo>

Amazon.com Services LLC

2020.06 - 2025.10

Software Engineer @ Transactional Data Storage

- Develop and operate Amazon's largest internal NoSql database for e-commerce infra at 3PB scale.
 - Redesign and implement backup&recovery system utilizing native service SQS, ECS for dynamic scaling. Hardware cost is reduced by 50% and operational cost is reduced by 80%.
 - Lead design/impl for point-in-time-recovery utilizing AWS batch and glue job, supporting rollback for up to 7 days at 5min interval.
 - Lead design/impl of data deletion and compaction service for cost management. Operating cost is reduced by 30% max for single cluster.
 - Lead peak day operational campaign that serves **7MM tps** traffic for biggest NA cluster with 99.999% availability.
- Design and develop **record-level regional-flexible** SQL-based OLTP database system with high consistency and availability at large scale
 - Improve Paxos algorithm to cope with multi-record, multi-version transaction use-case
 - Design and deliver core components of control plane e.g. centralized metadata service, capacity management service, workflow orchestration service and feedback control loop system.
 - Design and develop region flexible feature that enables per-table record-level data movement with on-fly transactions. This greatly enables Amazon to move data in a seamlessly way between regions to meet data compliance.
 - Design and develop reconfiguration algorithm/workflow that dynamically re-balance the node to scale/descale. It eliminates the campaign to manually reshard the cluster to scaling up thus improved better user experience for peak days.
 - Design and develop backup and recovery system utilizing S3 as network drive and garbage collecting learned value into blocked files. Compared with traditional database backup is at almost real time and recovery speed is reduced by 90%.
- Develop global infrastructure management platform on top of AWS to host large-scale stateful services for e-commerce
 - Utilize EKS for auto scaling, failure mitigation etc. to separate software and infrastructure maintenance.
 - Utilize VPC to provide flat global network to support region flexible goal.
- Lead campaign and coordinate stake holders for data movement campaign to new data centers to avoid energy shortage. Total 15TB data is moved without degraded availability/consistency.

EDUCATION

University of California San Diego

2018.09 - 2020.03

Master of Science in Electrical and Computer Engineering

- GPA:3.71/4.0
- **Relevant courses:** motion planning and optimal control, robotics estimation, deep learning applications, statics learning, reinforcement learning.
- Academic experience: design a 7-Dof lung biopsy surgical robotics for CT scan, mainly in charge of software simulation and kinematics control algorithm.

Wuhan University of Technology

2014.09 - 2018.06

Bachelor of Science in Mechanical Engineering

- In 2016 went to UC Davis for a one-year exchange program, during which joint FSAE team(main duty is vehicle dynamics and stability analysis)
- **Honors/Awards:** Outstanding graduates(2018), Outstanding Extracurricular activities individual(2018)