Kirkland
USA
+1 (425)-615-3057
xudifsd@gmail.com
github.com/xudifsd

Di Xu

Skills

- **Languages**: C++, C#, Python, Java, Clojure
- **Technologies**: Linux, Kafka, Kubernetes, Raft, MySQL

Key Achievements

- **Service design and team lead**: Designed and implemented Feature Extraction Library(FEL) and Content Builder(CB). Led team adding calculation logic and type supports to FEL, adding Python tool for configuring FEL. Led team optimizing CB and improving service relibilties.
- **System Reliability**: Re-architected Matrix master server by changing data storage from MySQL to Raft, improved system availability from 99.9% to 99.95% and shortened failover time from 30min to 1min.
- **Cross team collaboration**: Collaborated with Tally core team on optimizing read path. Led user migration to FEL and CB.

Working Experience

Microsoft Corporation, Principal Software Engineer, Redmond

2025/09 – now

Recommandation platform

- Designed and implemented Content Builder based on **Kafka**.
  - The service is strategically important to the team, providing capabilities for injection time feature calculations.
  - Improved team agility and reduced manual config deploying.
  - Queue based solution makes handling concurrency more elegantly. Previously we used CAS to ensure concurrent writes will not result lost data, now kafka ensures update to the same key will be handled sequentially.
- Implemented a service for supporting ranking documents for LLM Grounding.

Microsoft Corporation, Sr. Software Engineer, Redmond

2021/09 – 2025/09

Recommandation platform

- Designed and implemented Feature Extraction Library in **C++**. Led the migration of all C# feature logic to the library, resulting in a 2ms reduction in 95th percentile E2E latency and a 15% decrease in CPU usage relating to feature calculations.
- Optimized read path for Tally counting service
  - Reduced 250ms out of 1s for DB keys preparations.
  - Reduced cache entry filling latency from 1.2ms to 0.6ms by changing cache structure.
  - Provided a breakdown on cache memory usage for each tally counter for DS, driving some low ROI counters retirement.

Microsoft Corporation, Software Engineer II, Bellevue

2019/12 – 2021/09

Deep Learning Training Service

- Optimized Job-Manager, reducing 95th percentile job creation time from 400s to 46s.
- Optimized init process, reducing job initialization time from 45s to 2s.

Recommandation platform

- Designed and implemented service downgrade capability preventing service break down under heavy load, improved availabilities in this scenario.

Microsoft(China) limited, Software Engineer II, Beijing

2018/06 OpenAI platform that provide complete **AI** model training and **resource management** capabilities
2019/12
- Owner of monitoring subsystem, implemented a careful designed structure to minimize the impact of hanging process.
- Designed and implemented PAI runtime, which is the contract between user's job and training platform. Unified runtime environment of hadoop based backend and **kubernetes** based backend.

Collaborated cross-functionally with the Bellevue team to build Deep Learning Training Service.
- Helped unify monitoring system of DLWorkspace with OpenAI, implement many monitoring features required by DLTS.
- Refactored Job-Manager which is the backbone of DLTS, leveraged knowledge gained in OpenPAI to make Job-Manager more scalable and efficient.
- Used **Kusto**'s data to implement a physical machine topology aware scheduling, improved training efficiency.

Baidu Inc., Software Engineer, Beijing

2016/03 Matrix platform, a **Borg**/**Kubernete** like distributed cluster management system
2018/06
- Re-architected master server: changed underlaying data storage from **MySQL** to **Raft**.
  - Improved system availability from 99.9% to 99.95%.
  - Shorten failover time from 30min to 1min.
  - Collaborated with Test team in designing test plan for functionalities and robustness.
  - Collaborated with SRE team in drafting failover and disaster recovery plan.
- Evaluated several resource allocation algorithms using allocator-simulator and introduced some into the system.
- Designed and implemented serveral features requested by PaaS and users, including but not limited to disk selection, host selection and resource isolation.

Alibaba Cloud Computing (Aliyun) Inc., R & D Engineer Intern, Beijing

2015/05 Pangu system, a **GFS** like distributed storage system
2015/09
- Collected & analysed usage data from production env, generate report to persuad users to use system more effectively, resulting in 50% less disk usage.
- Analysed & restricted policy of list & read APIs, improved QPS from 110k to 150k.

BearyInnovative Inc., Softerware Engineer, Beijing

2014/08 BearyChat, a Slack like product
2015/05
- Designed backend topology, re-architected service from single node to distributed nodes, eliminated the risk of timestamp conflicting & single point of failure.
- Implemented 4 critical & many other important product features in web API server written in **Clojure**.

AdMaster Inc., R & D Engineer Intern, Beijing

2013/04 Data collecting system
2013/09
- Implemented 10 data processing algorithms in cascading, answered how many UV & PV per ad and characteristics of the viewers. These data will be presented to users directly.
- Maintained & optimized data collecting module written in **Python**, speeded it up by 13%.

Project Experience

Google Summer of Code - Typed Clojure

2014/05 Added 2 function type annotations to **Typed Clojure**. Made the type system more sound.
2014/08

- Came up & implemented a design much more expressive than mentor's.
- Improved core library annotation coverage by 7%, made Typed Clojure more user friendly.

Java Symbolic Executor targeting Android Apps

2013/10 Used by Android analysis team in Tencent Inc. Sponsored by CCF-Tencent Open Fund.
2014/05

- Designed & implemented core abstraction, enhanced robustness & readability.
- Optimized the program before shipping to the users, reduced its memory usage by 65%.

Education

2013/09 M.S. in **Softerware Design**, **University of Science and Technology of China**
2016/03

2009/09 B.S. in **Computer Science**, Dalian Polytechnic University
2013/06