

Qian LIN

Data Platform Team, ByteDance, Singapore

qian.lin@bytedance.com • (+65) 8359-0479 • <https://linqian.me>

► *GitHub*: <https://github.com/streamjoin> ► *SlideShare*: <https://www.slideshare.net/lqmike>

EDUCATION	Doctor of Philosophy (Ph.D.) in Computer Science <i>National University of Singapore (NUS)</i> , Singapore Research Areas: database systems, real-time data analytics, parallel and distributed computing	Aug 2012 – Nov 2017
	Master of Science (MSc) in Software Engineering <i>Shanghai Jiao Tong University (SJTU)</i> , Shanghai, China Research Areas: virtualization technology, operating system, cloud computing	Sep 2008 – Mar 2011
	Bachelor of Science (BSc) in Optical Information Science and Technology <i>South China University of Technology (SCUT)</i> , Guangzhou, China Cumulative GPA: 86.87 / 100; also attained a minor degree in Computer Science and Technology	Sep 2004 – Jul 2008
WORK EXPERIENCE	Software Engineer at <i>ByteDance</i> , Singapore Led the research and development of a high-performance real-time data processing system which can process tens of millions of streaming events per second. Moreover, my team further extended the system to support the SQL semantics, thereby providing a more friendly programming interface for data analysts.	Sep 2020 – Present
	Senior Research Fellow at <i>School of Computing, NUS</i> , Singapore Led research projects on distributed databases, cloud computing and artificial intelligence, and manage the development of related systems. Coordinated joint research with industrial partners in the projects. I publish papers in premier venues, and serve as program committee member in top-ranked conferences.	Jan 2019 – Sep 2020
	Postdoctoral Research Fellow at <i>School of Computing, NUS</i> , Singapore Led research projects on blockchain, distributed databases, deep learning and artificial intelligence, as well as developing core components of the related systems. Reviewed papers as an external reviewer.	Dec 2017 – Dec 2018
	Research Assistant at <i>School of Computing, NUS</i> , Singapore Participated in the development of (sub-)systems and tools of distributed database, data curation, deep learning and artificial intelligence. Published research papers in premier conferences and journals.	Apr 2016 – Nov 2017
	Intern at <i>Tencent</i> , Guangzhou, China Participated in the discussions on WeChat backend technologies, particularly on the high-availability storage system and the overload management for large-scale microservice architecture.	Jan – Jun 2017
	Software Engineer at <i>IBM China Development Lab</i> , Shanghai, China Participated in the product development of IBM System X server management tools.	Apr 2011 – Jul 2012
	Intern at <i>Tencent</i> , Shanghai, China Participated in the training program of massively multiplayer online game server development.	Jun – Sep 2010
RESEARCH EXPERIENCE	Fork Intelligence: Git for Data + AI Facilitating Services , NUS The platform aims to provide several AI oriented services in the cloud, including collaborative data cleansing, privacy-preserving machine learning, and quality control for online prediction. ForkBase is adopted and extended in the platform to support data management and access control in AI development activities. I lead the team to design and implement the platform towards commercialization.	Nov 2018 – Sep 2020
	ForkBase: a Universal Storage System for Forkable Applications , NUS The system is designed to unify and add values to many classes of today's applications with high-level semantics, high flexibility and performance. Specifically, it is designed as the storage support for applications that can benefit from data versioning and branching. My job focuses on developing the system kernel, including the content-addressable physical storage, management of data versioning and branching, key-value store, columnar relational store, clustered document store, application programming interface (API) and command-line interface (CLI).	Apr 2016 – Sep 2020

Foodlg: a Food Journaling, Nutrition Tracking and Analysis App, NUS Oct 2015 – Sep 2020
This preventive care app (<http://foodlg.com>) is designed for handy journaling of daily meals to help achieve a well-balanced diet, by simply taking pictures of what you have eaten. It is powered by deep learning that learns the food types from massive annotated images. In addition, a food-health knowledge base is built at the backend to link the food information with calories, ingredient, nutrition facts, etc. My job focuses on developing the image-based food recognition engine and designing the mobile app. I am also responsible for coordinating the cooperation with hospitals and research institutes on the use of this app as well as the adoption of its powering technology.

Rapiki: Effortless Cluster Computing, NUS Mar 2017 – Aug 2017
This is a framework designed for ease of carrying out embarrassingly parallel data processing at scale. It facilitates data scientists who have limited expertise in distributed computing to fully make use of cluster computing, with just minimal effort in configuration. I created this framework for the multimedia and computer vision research labs in NUS, as well as making it generic for coordinating independent tasks to run in parallel.

Trajectory Analysis Using Cell Phone Signal Data, NUS & StarHub Aug 2015 – Mar 2016
Nowadays mobile data network has become the main transmission media that enables people to use various kinds of apps on their smart phones at all times and places, especially when staying outside. Using cell phone signal (e.g., 3G/4G) data to do real-time trajectory analysis brings new insights of traffic dynamics, comparing with the traditional GPS-based methods. My job focuses on building a comprehensive tool of data cleansing and integration for pre-processing the massive cell phone signal data, and further mapping the location-based signal data to the roads to initiate the pipeline of trajectory analysis.

epiC: an Elastic, Power-aware, Data-intensive Cloud Platform, NUS Dec 2014 – Apr 2016
The objectives are to design and implement an efficient multi-tenancy cloud system for supporting high throughput low latency transactions and high performance reliable query processing, with online and interactive analytics capability. My job focuses on designing a stream processing engine to support real-time data analytics.

LogBase: a Distributed Log-structured Data Management System, NUS Aug 2013 – Feb 2015
The system adopts the log-only storage to handle high append and write load, such as sensor information processing. Indexing, transaction management and query processing are the key issues being investigated. My job focuses on designing an in-memory transaction processing engine based on data ownership association.

TEACHING EXPERIENCE

Teaching Assistant of CS6203 Advanced Topics in Database Systems, NUS 2017 – 2019
This seminar-style course covers the topics in database management systems with current research and industrial interests and importance. My responsibilities included developing course project assignments, coaching for students, and grading students' performance of paper presentation and class-room participation.

Lecturer of Introduction to Java Programming, NUS Dec 2015
This was a crash course in Java programming for the students who took the programme of master of science in business analytics. Apart from an introductory overview of the Java programming language, the lectures I delivered also covered the major principles of object-oriented programming.

Teaching Assistant of CS3223 Database Systems Implementation, NUS 2014
This course provided an in-depth study of the concepts and implementation issues related to database management systems. It covered topics of storage management, data indexing, query processing and optimization, transaction management, concurrency control, and failure recovery. My responsibilities included conducting weekly tutorials, coaching for students, and grading assignments.

Teaching Assistant of Linux and Open Source, SJTU 2009, 2010
This was an introductory course in Linux, providing the knowledge and hands-on skills of using Linux for development. It also covered the fundamentals of Linux internals such as system call, process management, virtual memory, I/O device driver, and file system. My responsibilities included developing lab assignments, preparing for reading lists, conducting coaching session, and grading assignments.

PROFESSIONAL SERVICES

Program committee member of **SIGMOD (2021)**, **ICDCS (2020)**, **VLDB (2019)**
Invited reviewer for **IEEE Transactions on Knowledge and Data Engineering (TKDE)**
Invited reviewer for **ACM Transactions on Database Systems (TODS)**
Invited reviewer for **IEEE Transactions on Parallel and Distributed Systems (TPDS)**
Invited reviewer for **Journal of Parallel and Distributed Computing (JPDC)**
Invited reviewer for **IEEE Transactions on Cloud Computing (TCC)**
Invited reviewer for **The VLDB Journal (VLDBJ)**
Invited reviewer for **ACM SIGMOD Record**
Invited reviewer for **World Wide Web Journal (WWWJ)**
Invited reviewer for **Journal of Distributed and Parallel Databases (DAPD)**
Invited reviewer for **IEEE Access**
Invited reviewer for **ACM Transactions on Internet Technology (TOIT)**
Invited reviewer for **ACM Transactions on Data Science (TDS)**
Invited reviewer for **Journal of Tsinghua Science and Technology (TST)**
Invited reviewer for **Journal of Big Data Research**

SELECTED PUBLICATIONS

P. Zhou, Q. Lin, D. Loghin, B. C. Ooi, Y. Wu, and H. Yu. Communication-efficient decentralized machine learning over heterogeneous networks. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE)*, 2021.

Z. Luo, S. H. Yeung, M. Zhang, K. Zheng, L. Zhu, G. Chen, F. Fan, Q. Lin, K. Y. Ngiam, and B. C. Ooi. MLCask: Efficient Management of Component Evolution in Collaborative Data Analytics Pipelines. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE)*, 2021.

P. Ruan, T. T. A. Dinh, D. Loghin, M. Zhang, G. Chen, Q. Lin, and B. C. Ooi. Blockchains vs. Distributed Databases: Dichotomy and Fusion. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, 2021.

Q. Lin, K. Yang, T. T. A. Dinh, Q. Cai, G. Chen, B. C. Ooi, P. Ruan, S. Wang, Z. Xie, M. Zhang, and O. Vandans. ForkBase: immutable, tamper-evident storage substrate for branchable applications (demo). In *Proceedings of the IEEE International Conference on Data Engineering (ICDE)*, 2020.

C. Song, Q. Lin, G. Ling, Z. Zhang, H. Chen, J. Liao, and C. Chen. LoCEC: local community-based edge classification in large online social networks. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE)*, 2020.

D. Loghin, S. Cai, G. Chen, T. T. A. Dinh, F. Fan, Q. Lin, J. Ng, B. C. Ooi, X. Sun, Q.-T. Ta, W. Wang, X. Xiao, Y. Yang, M. Zhang, and Z. Zhang. The disruptions of 5G on data-driven technologies and applications. In *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 2020.

P. Ruan, G. Chen, T. T. A. Dinh, Q. Lin, and B. C. Ooi. Fine-grained, secure and efficient data provenance on blockchain systems. In *Proceedings of the VLDB Endowment*, 12(9):975–988, 2019. **[Best Paper Award of VLDB 2019; Selectively reviewed by the Morning Paper on 16 Sep 2019]**

H. Dang, T. T. A. Dinh, D. Loghin, E.-C. Chang, Q. Lin, and B. C. Ooi. Towards scaling blockchain systems via sharding. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, 2019.

Q. Lin, G. Chen, and M. Zhang. On the design of adaptive and speculative concurrency control in distributed databases. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE)*, 2018.

S. Wang, T. T. A. Dinh, Q. Lin, Z. Xie, M. Zhang, Q. Cai, G. Chen, W. Fu, B. C. Ooi, and P. Ruan. Fork-Base: an efficient storage engine for blockchain and forkable applications. In *Proceedings of the VLDB Endowment*, 11(10):1137–1150, 2018. **[Selectively reviewed by the Morning Paper on 1 Jun 2018]**

C. Yao, M. Zhang, Q. Lin, B. C. Ooi, and J. Xu. Scaling distributed transaction processing and recovery based on dependency logging. In *the VLDB Journal*, 2018.

H. Zhou, M. Chen, Q. Lin, Y. Wang, X. She, S. Liu, R. Gu, B. C. Ooi and J. Yang. Overload control for scaling WeChat microservices. In *Proceedings of the ACM Symposium on Cloud Computing (SoCC)*, 2018. **[Selectively reviewed by the Morning Paper on 16 Nov 2018]**

Q. Lin. Online data processing at scale. *PhD Thesis*, NUS, 2017.

J. Zheng, Q. Lin, J. Xu, C. Wei, C. Zeng, P. Yang, and Y. Zhang. PaxosStore: high-availability storage made practical in WeChat. In *Proceedings of the VLDB Endowment*, 10(12):1730–1741, 2017.

Q. Lin, P. Chang, G. Chen, B. C. Ooi, K.-L. Tan, and Z. Wang. Towards a non-2PC transaction management in distributed database systems. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, 2016.

C. Yao, D. Agrawal, G. Chen, Q. Lin, B. C. Ooi, W.-F. Wong, and M. Zhang. Exploiting single-threaded model in multi-core in-memory systems. In *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 28(10):2635–2650, 2016.

Q. Lin, B. C. Ooi, Z. Wang, and C. Yu. Scalable distributed stream join processing. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, 2015.

D. Basu, Q. Lin, W. Chen, H. T. Vo, Z. Yuan, P. Senellart, and S. Bressan. Cost-model oblivious database tuning with reinforcement learning. In *Proceedings of the International Conference on Database and Expert Systems Applications (DEXA)*, 2015.

C. Zi, C. Zhang, Q. Lin, Z. Qi, and S. Gao. Suspend-to-PCM: a new power-aware strategy for operating system’s rapid suspend and resume. In *Emerging Technologies for Information Systems, Computing and Management*, 667–674, 2013.

Q. Lin, Z. Qi, J. Wu, Y. Dong, and H. Guan. Optimizing virtual machines using hybrid virtualization. In *Journal of Systems and Software (JSS)*, 85(11):2593–2603, 2012.

Q. Lin, M. Xia, M. Yu, P. Yu, M. Zhu, S. Gao, Z. Qi, K. Chen, and H. Guan. SPAD: software protection through anti-debugging using hardware virtualization. In *Proceedings of the ACM Symposium on Applied Computing (SAC)*, 2011.

M. Yu, Q. Lin, B. Li, Z. Qi, and H. Guan. Vis: virtualization enhanced live acquisition for native system. In *Proceedings of the ACM Asia-Pacific Workshop on Systems (APSys)*, 2011.

Q. Lin, and Y. Zhu. Software design of MIFARE smart IC card read-write device based on MCS-51. In *Electronic Engineer*, 33(11):54–57, 2007.

AWARDS

ACM SIGMOD Research Highlight Award	2020
Best Paper Award, VLDB	2019
Dean’s Graduate Research Excellence Award, NUS	2016
Research Achievement Award, NUS	2015
Research Scholarship, NUS	2012 – 2016
Academic Scholarship, SJTU	2009, 2010
Triple-A Student Scholarship, SCUT	2005, 2006, 2007