

# Le Xu

Phone: 217-721-4989

Email: contact@lexu.space

<https://lexu.space>

---

## Research Interest

Distributed Systems; Cloud Computing; Resource Management; Adaptivity; System Elasticity; Performance Analysis

## Education

- 08/2015–12/2021 University of Illinois at Urbana-Champaign  
Ph.D. in Computer Science (advised by Indranil Gupta)  
Thesis: “Elastic techniques to handle dynamism in real-time data processing systems.”
- 08/2013–05/2015 University of Illinois at Urbana-Champaign  
M.S in Computer Science (advised by Indranil Gupta)  
Thesis: “Stela: on-demand elasticity in distributed data stream processing systems”
- 08/2009–05/2013 University of Illinois at Urbana-Champaign  
B.S in Math and Computer Science)

## Positions

- 04/2024–Present Computing Infrastructure Lab, Bytedance Inc.  
Researcher/Software Engineer  
*Description: Building next-generation real-time serving services for generative models.*
- 09/2021–04/2024 University of Texas at Austin  
Postdoctoral Researcher (PI: Aditya Akella)  
*Description: Supporting fine-grained resource provisioning for modern real-time data processing systems.*

## Publication

- Peter Schafhalter, Sukrit Kalra, Le Xu, Joseph E. Gonzalez, and Ion Stoica. “Leveraging Cloud Computing to Make Autonomous Vehicles Safer.” IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2023
- Li Su, Xiaoming Qin, Zichao Zhang, Rui Yang, Le Xu, Indranil Gupta, Wenyuan Yu, Kai Zeng, Jingren Zhou “Banyan: A Scoped Dataflow Engine for Graph Query Service.” Proceedings of the VLDB Endowment (VLDB), 2022.
- Li Su, Xiaoming Qin, Zichao Zhang, Rui Yang, Le Xu, Indranil Gupta, Wenyuan Yu, Kai Zeng, Jingren Zhou “Banyan: A Scoped Dataflow Engine for Graph Query Service.” Proceedings of the VLDB Endowment (VLDB), 2022.
- Le Xu, Shivaram Venkataraman, Indranil Gupta, Luo Mai, Rahul Potharaju. “Move Fast and Meet Deadlines: Fine-grained Real-time Stream Processing with Cameo” USENIX Conference on Networked Systems Design and Implementation (NSDI), 2021. (*acceptance rate 16.6%*)

- Yunhui Long, Le Xu, Carl Gunter “A Hypothesis Testing Approach to Sharing Logs with Confidence” ACM Conference on Data and Application Security and Privacy (CO-DASPY), 2020.
- Faria Kalim, Le Xu, Sharanya Bathey, Richa Meherwal, Indranil Gupta. “Henge: Intent-driven Multi-Tenant Stream Processing” Symposium of Cloud Computing” Proceedings of the ACM Symposium on Cloud Computing (SoCC), 2018. (*acceptance rate 24.3%*)
- Luo Mai, Kai Zeng, Rahul Potharaju, Le Xu, Steve Suh, Shivaram Venkataraman, Paolo Costa, Terry Kim, Saravanan Muthukrishnan, Vamsi Kuppa, Sudheer Dhulipalla, Sriram Rao. “Chi: a scalable and programmable control plane for distributed stream processing systems.” Proceedings of the VLDB Endowment (VLDB), 2018.
- Mainak Ghosh, Ashwini Raina, Le Xu, Xiaoyao Qian, Indranil Gupta, Himanshu Gupta. “Popular is Cheaper: Curtailing Memory Costs in Interactive Analytics Engines.” Proceedings of the 2018 European Conference on Computer Systems, (EuroSys), 2018. (*acceptance rate 16.5%*)
- Mainak Ghosh, Le Xu, Indranil Gupta. “Resource Management: Performance Assuredness in Distributed Cloud Computing via Online Reconfigurations” Assured Cloud Computing. John Wiley & Sons, 2018.
- Mijung Kim, Jun Li, Haris Volos, Manish Marwah, Alexander Ulanov, Kimberly Keeton, Lucy Cherkasova, Le Xu, Pradeep Fernando. “Sparkle: Optimizing Spark for Large Memory Machines and Analytics” 2017 ACM Symposium on Cloud Computing (SOCC), poster track. 2017.
- Le Xu, Boyang Peng, and Indranil Gupta. “Stela: Enabling Stream Processing Systems to Scale-in and Scale-out On-demand.” *IEEE International Conference on Cloud Engineering (IC2E)*. 2016. (*acceptance rate 21.9%*)
- Wenting Wang, Le Xu, and Indranil Gupta. “Scale Up vs. Scale Out in Cloud Storage and Graph Processing Systems.” *2015 IEEE International Conference on Cloud Engineering (IC2E)*. IEEE, 2015.

## Preprints

- Jiamin Li, Le Xu, Hong Xu, Aditya Akella. ”BlockLLM: Multi-tenant Finer-grained Serving for Large Language Models”, 2024
- Bodun Hu, Le Xu, Jeongyoon Moon, Neeraja J. Yadwadkar, Aditya Akella. ”MOSEL: Inference Serving Using Dynamic Modality Selection.”, 2023.
- Le Xu, Divyanshu Saxena, Neeraja J. Yadwadkar, Aditya Akella and Indranil Gupta. ”Dirigo: Self-scaling Stateful Actors For Serverless Real-time Data Processing.”, 2023.

## Industrial Experience

- 06/2019-08/2019 Alibaba Damo Academy (Data Analytics and Intelligence Lab)  
Building hierarchical actor-based framework for distributed graph querying service
- 05/2017-08/2017 Microsoft (Cloud and Information Services Lab)  
Building a control layer inside of a real-time stream processing engine for flexible and efficient online monitoring and re-configuration
- 05/2016-08/2016 Hewlett-Packard Labs (Software Analytics Group)  
Conducting Spark performance analysis for micro-benchmark and machine learning applications
- 05/2015-08/2015 Salesforce Inc. (ServiceCloud Performance)

Implemented Jmeter Tests suites for ServiceCloud application platform with synthetic data set

05/2014-08/2014 Yahoo! Sunnyvale (Cloud Services and Organization)

Implemented a server log-processing framework for server log retrieval and analysis. (Pig, Hive & Oozie)

## Teaching Experience

01/2015, 01/2016 Cloud Computing Concepts (Coursera) - Teaching Assistant

01/2016 Advanced Distributed Systems - Teaching Assistant

01/2015, 09/2019 Distributed System - Teaching Assistant

05/2018, 05/2020 Cloud Computing Capstone

## Awards and Honors

2021 2021 CRA/CCC Computing Innovation Fellows (CIFellows)

2020 2020 EECS Rising Stars

2020 Tapia 2020 Conference Scholarship

2016 David J. Kuck Outstanding M.S. Thesis Award

2015 Outstanding Teaching Assistant

Travel Grant: SOSP 2019, OSDI 2018, SoCC 2018, SOSP 2017, Grace Hopper Celebration 2015, IWCA 2015

## Program Committee

02/2023 2023 ACM/IFIP/USENIX International Middleware Conference

05/2023 2024 USENIX Symposium on Networked Systems Design and Implementation

June 20, 2024