

# Yongjoo Park

Thomas M. Siebel Center, Rm 2114  
201 North Goodwin Avenue  
Urbana, IL 61801-2302

Website: <https://yongjoopark.com>  
Email: [yongjoo@g.illinois.edu](mailto:yongjoo@g.illinois.edu)

## Research

Data management, Systems for data science and machine learning, Leading **CreateLab**

## Academic Positions

### University of Illinois at Urbana–Champaign

Assistant Professor, Department of Computer Science. Jan 2021 – Present

National Center for Supercomputing Applications (NCSA) Fellow. 2023 – Present

Faculty Affiliate, Informatics Programs, School of Information. 2022 – Present

Adjunct Assistant Professor (leap years), Department of Computer Science. Nov 2019 – Dec 2020

- During this period, I **co-founded Keebo, Inc., raising \$15M** from venture capital investors.

### University of Michigan, Ann Arbor

Research Fellow (postdoc), Computer Science and Engineering. Sep 2017 – Jul 2019

- *Principal Investigator*: Barzan Mozafari

## Education

### University of Michigan, Ann Arbor

Ph.D. in Computer Science and Engineering. Aug 2017

- *Advisors*: Michael Cafarella and Barzan Mozafari
- Dissertation is awarded 2018 **ACM SIGMOD Jim Gray Dissertation Award Runner-Up**

M.S. in Computer Science. Jun 2013

### Seoul National University

B.S. in Electrical Engineering. Feb 2009

## Professional Experience

### Keebo, Inc. (startup co-founded by me: \$23M raised from VC firms)

Chief Scientist. Sep 2022 – Present

- The company raised \$8M in December 2023 (Series A+)

Co-founder and CTO. Aug 2019 – Aug 2022

- I am one of the two co-founders. The company raised \$15M (Seed and Series A).
- Customers include Costco Travel, Allbirds, and Barstool Sports.

### University of Michigan, Ann Arbor

Graduate Student Research Assistant. Sep 2012 – Apr 2017

### **Amazon.com, Seattle**

Software Engineer Intern. May 2014 – Aug 2014

### **University of Michigan, Ann Arbor**

Graduate Student Instructor. Jan 2012 – Apr 2012

### **Webcash, Seoul**

Software Engineer. Dec 2008 – May 2011

### **Seoul National University**

Research Assistant. Jun 2007 – Jan 2008

## Honors & Awards

**ACM SIGMOD 2023 Best Artifact Award Honorable Mention.** May 2024

UIUC Teachers Ranked as Excellent (CS 511 Advanced Data Management): Fall 2022, Fall 2023

UIUC Engineering Council Outstanding Advising Award. Feb 2021

**ACM SIGMOD Jim Gray Dissertation Award Honorable Mention.** Jun 2018

ACM SIGMOD Student Travel Award. May 2017

University of Michigan Rackham Travel Grant. Jan 2017

Kwanjeong Graduate Study Fellowship (for Ph.D.) \$100,000. 2013 – 2017

Jeongsong Graduate Study Fellowship (for M.S.) \$55,000. 2013 – 2017

Korean National Science Scholarship \$20,000. 2004

## Funding (total: \$1.3M; share: \$850K)

ACCESS Allocations 1.5M credits (which amounts to \$11,858). 2024

Microsoft Azure Cloud Credit, \$10,000. 2023

PI, **NSF III Medium Award (# 2312561)** titled “CARE: Interactive Systems for Scalable, Causal Data Science” \$1,216,000 (share: \$748,000). 2023 – 2027

PI, **National Center for Computing Applications Fellow** titled “TheVault: Accelerating Scientific Discovery with Reliable and Reproducible Machine Learning” \$25K (share: \$25K). 2023 – 2024

Microsoft Azure Cloud Credit, \$16,000. 2022

Google Cloud Credits, \$5,000. 2021

Microsoft Azure Cloud Credit, \$30,000. 2020

Google Cloud Credits, \$10,000. 2020

## Publication

### **Peer-Reviewed Papers**

*All full-length research papers unless noted otherwise (e.g., demo, short)*

21. Zhaoheng Li, Supawit Chockchowwat, Hanxi Fang, Ribhav Sahu, Sumay Thakurdesai, Kantanat Pridaphatrakun, **Yongjoo Park**. “Demonstration of ElasticNotebook: Migrating Live Computational Notebook States.” SIGMOD’24: Proceedings of the ACM on Management of Data, Santiago, Chile, 2024. (*demo*) (Acceptance Rate Unavailable; Average Rate: 20%)
20. Supawit Chockchowwat, Wenjie Liu, and **Yongjoo Park**. “AirIndex: Versatile Index Tuning Through Data and Storage.” SIGMOD’24: Proceedings of the ACM on Management of Data, Santiago, Chile, 2024. (Acceptance Rate Unavailable; Average Rate: 20%)
19. Zhaoheng Li, Pranav Gor, Rahul Prabhu, Hui Yu, Yuzhou Mao, and **Yongjoo Park**. “ElasticNotebook: Enabling Live Migration for Computational Notebooks.” PVLDB’23: Proceedings of the VLDB Endowment 17.2 (2023): 119-133. (Acceptance Rate Unavailable, Average Rate: 18.6%)
18. Ipoom Jeong, Jiaqi Lou, Yongseok Son, **Yongjoo Park**, Yifan Yuan, and Nam Sung Kim. “LADIO: Leakage-Aware Direct I/O for I/O-Intensive Workloads.” CAL’23: IEEE Computer Architecture Letters, 2023. (*short*) (Average Acceptance Rate: 20%)
17. Supawit Chockchowwat, Zhaoheng Li, and **Yongjoo Park**. “Transactional Python for Durable Machine Learning: Vision, Challenges, and Feasibility.” DEEM’23: In Proceedings of the Seventh Workshop on Data Management for End-to-End Machine Learning, Seattle, USA, 2023.
16. Barzan Mozafari, Radu Alexandru Burcuta, Alan Cabrera, Andrei Constantin, Derek Francis, David Grömling, Alekh Jindal, Maciej Konkolowicz, Valentin Marian Spac, **Yongjoo Park**, Russell Razo Carranzo, Nicholas Richardson, Abhishek Roy, Aayushi Srivastava, Isha Tarte, Brian Westphal, and Chi Zhang. “Making Data Clouds Smarter at Keebo: Automated Warehouse Optimization using Data Learning.” SIGMOD’23: In Companion of the 2023 International Conference on Management of Data, 2023. (*industry track*)
15. Nikhil Sheoran, Supawit Chockchowwat, Arav Chheda, Suwen Wang, Riya Verma, and **Yongjoo Park**. “A Step Toward Deep Online Aggregation.” SIGMOD’23: Proceedings of the ACM on Management of Data, Seattle, USA, 2023. (Acceptance Rate: 27.2%)  
*Awarded “Artifacts Available”, “Artifacts Evaluated”, and “Results Reproduced” badges*  
**SIGMOD’23 Best Artifact Award Honorable Mention**
14. Zhaoheng Li, Xinyu Pi, and **Yongjoo Park**. “S/C: Speeding up Data Materialization with Bounded Memory” ICDE’23: IEEE 39th International Conference on Data Engineering, Anaheim, USA, 2023. (Acceptance Rate: 30.6%)
13. Supawit Chockchowwat, Wenjie Liu, and **Yongjoo Park**. “Automatically Finding Optimal Index Structure.” AIDB’22: 4th International Workshop on Applied AI for Database Systems and Applications, Sydney, Australia, 2022.
12. Sophia Yang, **Yongjoo Park**, and Abdussalam Alawini. FIE’22: “The Effects of Teaching Modality on Collaborative Learning: A Controlled Study.” The Frontiers in Education, Uppsala, Sweden, 2022. (Acceptance Rate: 55%)
11. Supawit Chockchowwat, Chaitanya Sood, and **Yongjoo Park**. “Airphant: Cloud-oriented Document

Indexing.” ICDE’22: IEEE 38th International Conference on Data Engineering, Kuala Lumpur, Malaysia, 2022. (Acceptance Rate: 27%)

10. Johes Bater, **Yongjoo Park**, Xi He, Xiao Wang, and Jennie Rogers. “SAQE: Practical Privacy-Preserving Approximate Query Processing for Data Federations.” Proceedings of the VLDB Endowment 13.12 (2020): 2691-2705. (Acceptance Rate: 24.8%)
9. **Yongjoo Park**, Shucheng Zhang, and Barzan Mozafari. “QuickSel: Quick Selectivity Learning with Mixture Models.” SIGMOD’20: ACM SIGMOD/PODS International Conference on Management of Data, Portland, OR, USA, 2020. (Acceptance Rate: 26.9%)
8. **Yongjoo Park**, Jingyi Qing, Xiaoyang Shen, and Barzan Mozafari. “BlinkML: Efficient Maximum Likelihood Estimation with Probabilistic Guarantees.” SIGMOD’19: ACM SIGMOD/PODS International Conference on Management of Data, Amsterdam, The Netherlands, 2019. (Acceptance Rate: 20.5%)
7. **Yongjoo Park**, Barzan Mozafari, Joseph Sorenson, and Junhao Wang “VerdictDB: Universalizing Approximate Query Processing.” SIGMOD’18: ACM SIGMOD/PODS International Conference on Management of Data, Houston, TX, USA, 2018. (Acceptance Rate: 19.5%)
6. Wen He, **Yongjoo Park**, Idris Hanafi, Jacob Yatvitskiy, and Barzan Mozafari “Demonstration of VerdictDB, the Platform-Independent AQP System.” ACM SIGMOD/PODS International Conference on Management of Data, Houston, TX, USA, 2018. (*demo*)
5. **Yongjoo Park**, Amhad Shahab Tajik, Michael Cafarella, and Barzan Mozafari. “Database Learning: Toward a Database System that Becomes Smarter Over Time.” SIGMOD’17: ACM SIGMOD/PODS International Conference on Management of Data, Chicago, IL, USA, 2017. (Acceptance Rate: 19.6%)
4. **Yongjoo Park**. “Active Database Learning.” CIDR’17: The biennial Conference on Innovative Data Systems Research, Chaminade, CA, USA, 2017. (*abstract*)
3. **Yongjoo Park**, Michael Cafarella, and Barzan Mozafari. “Visualization-Aware Sampling for Very Large Databases.” ICDE’16: IEEE 32nd International Conference on Data Engineering, Helsinki, Finland, 2016. (2016 Acceptance Rate Unavailable; Average Rate: 19.1%)
2. **Yongjoo Park**, Michael Cafarella, and Barzan Mozafari. “Neighbor-Sensitive Hashing.” PVLDB’15: Proceedings of the VLDB Endowment 9.3 (2015): 144-155 (Acceptance Rate: 35.5%)
1. Michael Anderson, Dolan Antenucci, Victor Bittorf, Matthew Burgess, Michael Cafarella, Arun Kumar, Feng Niu, **Yongjoo Park**, Christopher Ré, and Ce Zhang. “Brainwash: A Data System for Feature Engineering.” CIDR’13: The biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, 2013.

## Patents

2. Alekh Jindal, Barzan Mozafari, **Yongjoo Park**, David Wolfgang Grömling, Brian Westphal, and Alan D Cabrera. “Managed tuning for data clouds.” US Patent 11,693,857, 2023
1. Alekh Jindal, Barzan Mozafari, **Yongjoo Park**, Brian Westphal, Shi Qiao, Matthew Larsen, and Advait Abhay Dixit. “Platform agnostic query acceleration.” US Patent 11,567,936, 2023

## Dissertation

1. **Yongjoo Park.** “Fast Data Analytics by Learning.” Ph.D. Dissertation.  
*Awarded 2018 ACM SIGMOD Jim Gray Dissertation Award Runner-Up*

## Major Software Releases

### Kishu

The first system that can checkpoint data science workflow, enabling *time machine* for almost all interactive data science explorations. Designed to easily integrate into Jupyter, one of the most popular data science platforms. <https://github.com/illinoisdata/kishu>

### VerdictDB

The first database system that can accelerate analytical queries by orders of magnitude without data migration. I developed this system as part of PhD and postdoc research. The codebase is still used by my startup company Keebo. Tested and deployed at Walmart and Dunnhumby. <https://verdictdb.org/>

## Advising

### PhD Students

- Haocheng Xia. 2024 – Present
- Xinying Zheng. 2023 – Present, with Indy Gupta
- Dohyun Park. 2023 – Present
- Billy Li. 2021 – Present
- Supawit Chockchowwat. 2020 – Present

### MS Thesis Students

- Talika Gupta. 2024 – Present
- Hanxi Fang. 2023 – Present
- Raunak Shah. 2023 – Present
- Arthur Huang. 2023 – Present
- Shuning Zhang. 2023 – Present
- Nikhil Sheoran. 2021 – 2022, First employment: Databricks

### Undergraduate Students

- 2024–2025: Tia Kashyap (woman), Shreya Jangada (woman), Sumay Thakurdesai (man), Areet Sheth (man), Aditya Shreenivas Kunte (man), Kongning Lai (woman), Shubham Mital (man)
- 2023–2024: Zora Zhang (woman), Sumay Thakurdesai (man), Krishna Rama Rao (man), Kantanat Prisdaphatrakun (man), Ribhav Sahu (man), Areet Sheth (man)
- 2022–2023: Hui Yu (woman), Riya Verma (woman), Rahul Prabhu (man), Ribhav Sahu (man), Shriyan Gosavi (man)
- 2021–2022: Hui Yu (woman), Wenjie Liu (man), Arav Chheda (man), Suwen Wang (man), Sean Nam (man)

- 2020–2021: Anxue Chen (woman), Shuning Zhang (woman), Chaitanya Sood (man), Siyi Yu (man), Richard Deng (man), Xinlong Dai (man), Karthik Talluri (man), Venkatesh Dharmala (man), Tianying Zhu (man), Vineet Chinthakindi (man)
- University of Michigan, Ann Arbor (prior to joining Illinois): Shucheng Zhong 2018–2020; Wen He 2017–2018; Jingyi Qing 2017–2018; Xiaoyang Shen 2017–2018; Junhao Wang 2017–2018

## Teaching (all at UIUC)

Advanced Data Systems (CS 511), Fall 2024.

ML and Data Systems (CS 598 YP), Spring 2024

Advanced Data Systems (CS 511), Fall 2023. *Teachers Ranked as Excellent*

Advanced Data Systems (CS 511), Fall 2022. *Teachers Ranked as Excellent*

Database Management Systems (CS 411), Fall 2021

Advanced Data Systems (CS 511), Spring 2021

## Service

Program Committee, European Conference on Computer Systems (EuroSys) 2026

Research Track Program Committee, International Conference on Very Large Data Bases (VLDB) 2026

Program Committee, 29th International Conference on Extending Database Technology (EDBT) 2026

Research Track Program Committee, International Conference on Very Large Data Bases (VLDB) 2025

Reviewer, ACM CHI Conference on Human Factors in Computing Systems (CHI) late-breaking work 2025

Program Committee, Seventh Conference on Machine Learning and Systems (MLSys) 2025

PhD Dissertation Committee, Hongtai Cao, UIUC, 2024

Reviewer, International Journal on Very Large Data Bases (VLDBJ) 2024

CS CARES Committee, UIUC, 2024-2025

PhD Dissertation Committee, Beomyeol Jeon, UIUC, 2024

Reviewer, IEEE Transactions on Knowledge and Data Engineering (TKDE) 2024-04

Program Committee, ACM Student Research Competition Grand Finals 2024

NSF Proposal Review Panel, 2024

Program Committee, Workshop on Data Management for End-to-End Machine Learning (DEEM) 2024

Program Committee, Seventh Conference on Machine Learning and Systems (MLSys) 2024

Graduate Admissions Committee, Computer Science, UIUC, 2024

Reviewer, IEEE Transactions on Knowledge and Data Engineering (TKDE) 2023

Reviewer, International Journal on Very Large Data Bases (VLDBJ) 2023

Program Committee, ACM SIGMOD Student Research Competition 2023

Reviewer, Information Systems Journal 2023

Program Committee, ACM Student Research Competition Grand Finals 2023

Graduate Admissions Committee, Computer Science, UIUC, 2023

Program Committee, Sixth Conference on Machine Learning and Systems (MLSys) 2023

**Publicity Chair**, IEEE International Conference on Data Engineering (ICDE) 2023

Program Committee, ACM Symposium on Cloud Computing (SoCC) 2022  
 PhD Dissertation Committee, Amirhossein Aleyasen, UIUC, 2022  
 Program Committee, ACM Student Research Competition Grand Finals 2022  
**Co-chair**, SIGMOD Student Research Competition 2022  
 Program Committee, Fifth Conference on Machine Learning and Systems (MLSys) 2022  
 Faculty search committee, Computer Science, UIUC, 2022  
 Program Committee, ACM Symposium on Cloud Computing (SoCC) 2021  
 Graduate Admissions Committee, Computer Science, UIUC, 2021  
**Co-chair**, SIGMOD Student Research Competition 2021  
 Research Track Program Committee, International Conference on Very Large Data Bases (VLDB) 2021  
 Research Track Program Committee, International Conference on Very Large Data Bases (VLDB) 2020  
 Program Committee, International Conference on Management of Data (SIGMOD) 2020  
 Program Committee, SIGMOD Student Research Competition 2020  
 Reviewer, IEEE Transactions on Knowledge and Data Engineering (TKDE) 2019  
 Reviewer, International Journal on Very Large Data Bases (VLDBJ) 2019  
 Program Committee, ACM Symposium on Cloud Computing (SoCC) 2019  
 Reviewer, IEEE Transactions on Knowledge and Data Engineering (TKDE) 2018  
 Program Committee, aiDM workshop at SIGMOD 2018 (<http://www.aidm-conf.org/>)  
 Reviewer, International Conference on Management of Data (SIGMOD) 2018  
 Publicity Chair, ACAIA workshop 2017 (<http://dbgroup.eecs.umich.edu/acaia/>)  
 Reviewer, International Journal on Very Large Data Bases (VLDBJ) 2017  
 Organizer of

- University of Michigan Database Group meetings 2016, 2014
- MIDAS (Michigan Data Science) seminars 2014

## Talks

- “Undoable Jupyter with a Novel Data Management Layer,” Columbia University, November 2024
- “TheVault: Accelerating Scientific Discovery with Reliable and Reproducible Machine Learning,” NCSA Fellow Lightning Talk, National Center for Supercomputing Applications, October 2024
- Database Seminar, Cornell University, March 2023
- CS Colloquium, University of Illinois Urbana-Champaign, via Zoom, February 2021
- Dream Lab, University of Massachusetts Amherst, via Zoom, November 2020
- AWS User Group, Chicago, Nov 2019
- SIGMOD, Amsterdam, June 2019
- WAX workshop at FCRC, Phoenix, June 2019
- Criteo NABD conference, Ann Arbor, May 2019
- University of Texas, Austin, April 2019
- Penn State University, State College, April 2019
- Purdue University, West Lafayette, April 2019

- Northeastern University, Boston, March 2019
- University of Waterloo, March 2019
- Georgia Tech, Atlanta, March 2019
- University of Illinois, Urbana-Champaign, March 2019
- Microsoft Research, Redmond, February 2019
- Northwestern University, Redmond, February 2019
- Microsoft, Redmond, February 2019
- IBM Research, Almaden, February 2019
- SIGMOD, Houston, June 2018
- AVL (www.avl.com), Ann Arbor, April 2018
- Oracle BI Group, Redwood City, December 2017
- ACAIA workshop, San Jose, November 2017
- Oracle Database Group, Redwood City, November 2017
- Cloudera Impala Team, Palo Alto, November 2017
- Big Data Innovation Summit, Boston, September 2017
- New Tech Meetup, Ann Arbor, July 2017
- SIGMOD, Chicago, May 2017
- University of Michigan Software Group, Ann Arbor, May 2017
- Brown Database Group, Providence, March 2017
- Stanford InfoLab, Palo Alto, February 2017
- CIDR, Chaminade, California, January 2017
- **Midwest Big Data Opportunities and Challenges (MBDOC) Workshop**, Chicago, September 2016
- **North East Database Day (NEDB)**, Boston, January 2016
- ICDE, Helsinki, Finland, May 2016
- AVL (www.avl.com), Ann Arbor, April 2016
- VLDB, New Delhi, India, September 2016
- VSM workshop at ICCV, Santiago, Chile, December 2015