

# Yongjoo Park

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

Website: <https://yongjoopark.com>  
Email: [yongjoo@illinois.edu](mailto:yongjoo@illinois.edu)  
Voice: +1 (734) 707-9206

RESEARCH	A.I. for data-intensive systems, Systems for analytics and machine learning	
ACADEMIC POSITIONS	<b>Assistant Professor</b> Department of Computer Science University of Illinois at Urbana–Champaign	From Jan 2021
	<b>Adjunct Assistant Professor</b> Department of Computer Science University of Illinois at Urbana–Champaign	Nov 2019–Dec 2020
	<b>Research Fellow</b> Computer Science and Engineering University of Michigan, Ann Arbor • <i>Principal Investigator:</i> Barzan Mozafari	Sep 2017–Jul 2019
EDUCATION	<b>Ph.D., Computer Science and Engineering</b> University of Michigan, Ann Arbor • <i>Advisors:</i> Michael Cafarella and Barzan Mozafari	Aug 2017
	<b>M.S., Computer Science</b> University of Michigan, Ann Arbor	Jun 2013
	<b>B.S., Electrical Engineering</b> Seoul National University (SNU)	Feb 2009
PROFESSIONAL EXPERIENCE	<b>Co-founder and CTO</b> Keebo, Inc.	Aug 2019–Present
	<b>Graduate Student Research Assistant</b> University of Michigan, Ann Arbor	Sep 2012–Apr 2017
	<b>Software Engineer Intern</b> Amazon.com, Seattle	May 2014–Aug 2014
	<b>Graduate Student Instructor</b> University of Michigan, Ann Arbor	Jan 2012–Apr 2012
	<b>Software Engineer</b> Webcash, Seoul	Dec 2008–May 2011
	<b>Research Assistant</b> Seoul National University, Seoul	June 2007–Jan 2008

AWARDS	2018 ACM SIGMOD Jim Gray Dissertation Award Runner-up	Jun 2018
	ACM SIGMOD Student Travel Award, \$900	May 2017
	Rackham Travel Grant, \$800	Jan 2017
	Kwanjeong Graduate Study Fellowship (for Ph.D.), \$100,000	2013–2017
	Jeongsong Graduate Study Fellowship (for Masters), \$55,000	2011–2013
	Korean National Science Scholarship, \$20,000	2004

- RESEARCH PROJECTS
- Machine Learning for Systems**
- Database Learning: DBL is the first approximate query processing system that can produce increasingly more accurate answers as it processes more queries.
  - Selectivity Learning: QuickSel is a selectivity estimation algorithm that becomes more accurate as it processes more queries.
- Systems for Fast Machine Learning**
- BlinkML is a *fast* ML system with *probabilistic quality guarantees*.
- Systems for Fast Data Analytics**
- SQL Analytics: VerdictDB is the first approximate query processing system that *can run on top of any SQL engines*.
  - Visualization: VAS is a sampling algorithm specialized for visualization scatter plots.
  - Image Search: NSH is a hashcode-based *k*-nearest neighbor algorithm.

- PUBLICATION
- Referred Conference Papers**
1. Johes Bater, **Yongjoo Park**, Xi He, Xiao Wang, Jennie Rogers  
**SAQE: Practical Privacy-Preserving Approximate Query Processing for Data Federations**  
**PVLDB 2020** (research): 46th International Conference on Very Large Data Bases, Tokyo, Japan (Online due to COVID-19), 2020.
  2. **Yongjoo Park**, Shucheng Zhang, Barzan Mozafari  
**QuickSel: Quick Selectivity Learning with Mixture Models**  
**SIGMOD'20** (research): ACM SIGMOD/PODS International Conference on Management of Data, Portland, OR, USA, 2020.
  3. **Yongjoo Park**, Jingyi Qing, Xiaoyang Shen, Barzan Mozafari  
**BlinkML: Efficient Maximum Likelihood Estimation with Probabilistic Guarantees**  
**SIGMOD'19** (research): ACM SIGMOD/PODS International Conference on Management of Data, Amsterdam, The Netherlands, 2019.
  4. **Yongjoo Park**, Barzan Mozafari, Joseph Sorenson, Junhao Wang  
**VerdictDB: Universalizing Approximate Query Processing**  
**SIGMOD'18** (research): ACM SIGMOD/PODS International Conference on Management of Data, Houston, TX, USA, 2018.

5. Wen He, **Yongjoo Park**, Idris Hanafi, Jacob Yatvitskiy, Barzan Mozafari  
**Demonstration of VerdictDB, the Platform-Independent AQP System**  
**SIGMOD'18** (demo): ACM SIGMOD/PODS International Conference on Management of Data, Houston, TX, USA, 2018.
6. **Yongjoo Park**, Amhad Shahab Tajik, Michael Cafarella, Barzan Mozafari  
**Database Learning: Toward a Database System that Becomes Smarter Over Time**  
**SIGMOD'17** (research): ACM SIGMOD/PODS International Conference on Management of Data, Chicago, IL, USA, 2017.  
*SIGMOD Travel Award*
7. **Yongjoo Park**  
**Active Database Learning**  
**CIDR'17** (abstract): The biennial Conference on Innovative Data Systems Research, Chaminade, CA, USA, 2017.
8. **Yongjoo Park**, Michael Cafarella, Barzan Mozafari  
**Visualization-Aware Sampling for Very Large Databases**  
**ICDE'16** (research): IEEE 32nd International Conference on Data Engineering, Helsinki, Finland, 2016.
9. **Yongjoo Park**, Michael Cafarella, Barzan Mozafari  
**Neighbor-Sensitive Hashing**  
**PVLDB'15** (research) for VLDB'16: 42nd International Conference on Very Large Data Bases, New Delhi, India, 2016.
10. Michael Anderson, Dolan Antenucci, Victor Bittorf, Matthew Burgess, Michael Cafarella, Arun Kumar, Feng Niu, **Yongjoo Park**, Christopher Ré, Ce Zhang  
**Brainwash: A Data System for Feature Engineering**  
**CIDR'13** (vision): The biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, 2013.

#### Thesis

11. Yongjoo Park  
**Fast Data Analytics by Learning**  
Ph.D. Dissertation  
*Awarded 2018 ACM SIGMOD Jim Gray Dissertation Award runner-up*

#### Non-Refereed Technical Reports (full versions to published papers)

12. Yongjoo Park, Jingyi Qing, Xiaoyang Shen, Barzan Mozafari  
**BlinkML: Efficient Maximum Likelihood Estimation with Probabilistic Guarantees**
13. Yongjoo Park, Barzan Mozafari, Joseph Sorenson, Junhao Wang  
**VerdictDB: Universalizing Approximate Query Processing**

14. Yongjoo Park, Amhad Shahab Tajik, Michael Cafarella, Barzan Mozafari  
**Database Learning: Toward a Database System that Becomes Smarter Over Time**
15. Yongjoo Park, Michael Cafarella, Barzan Mozafari  
**Neighbor-Sensitive Hashing**
16. Yongjoo Park, Michael Cafarella, Barzan Mozafari  
**Visualization-Aware Sampling for Very Large Databases**

TEACHING

**Advanced Database Management Systems (EECS 584)**  
Guest Lecturer, University of Michigan, Fall 2018

**Database Management Systems (EECS 484)**  
Guest Lecturer, University of Michigan, Winter 2018

**Web Databases and Information Systems (EECS 485)**  
Graduate Student Instructor, University of Michigan, Winter 2012

ADVISING

Supawit Chockchowwat (PhD., University of Illinois at Urbana-Champaign)	2020–Present
Chaitanya Sood (B.S., University of Illinois at Urbana-Champaign)	2020–Present
Shucheng Zhong (B.S. and M.S., University of Michigan, Ann Arbor)	2018–2020
Wen He (B.S., University of Michigan, Ann Arbor)	2017–2018
Jingyi Qing (B.S., University of Michigan, Ann Arbor)	2017–2018
Xiaoyang Shen (B.S., University of Michigan, Ann Arbor)	2017–2018
Joseph Sorenson (M.S., University of Michigan, Ann Arbor)	2017–2018
Junhao Wang (B.S., University of Michigan, Ann Arbor)	2017–2018

TALKS

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  
MBDOC, Chicago, September 2016  
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

SERVICE

Co-chair, SIGMOD 2021–2022 Student Research Competition  
Research Track Program Committee, VLDB 2021  
Research Track Program Committee, VLDB 2020  
Program Committee, SIGMOD 2020  
Program Committee, SIGMOD 2020 Student Research Competition  
Reviewer, TKDE 2019  
Reviewer, VLDB Journal 2019  
Program Committee, SoCC 2019

Reviewer, TKDE 2018

Program Committee, aiDM workshop at SIGMOD 2018 (<http://www.aidm-conf.org/>)

Reviewer, SIGMOD 2018

Publicity Chair, ACAIA workshop 2017 (<http://dbgroup.eecs.umich.edu/acaia/>)

Reviewer, VLDB Journal 2017

Organizer of

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