# ZHUOYUE ZHAO

72 S. Central Campus Drive, Rm 2780, Salt Lake City, UT, 84112
zyzhao [at] cs.utah.edu ⋄ http://www.cs.utah.edu/~zyzhao/

## EDUCATION

**University of Utah** August 2016 - Summer, 2021 (expected)
Ph.D. Student in Computer Science
School of Computing
Advisor: Prof. Feifei Li
GPA: 4.0/4.0
Thesis research: Approximate Query Proccessing via Random Sampling

**Shanghai Jiao Tong University** Sept. 2012 - July 2016
B.S.E. in Computer Science and Technology (SJTU ACM Class)
Zhiyuan College
Overall GPA: *3.96/4.30* Major GPA: *4.09/4.30* Ranking: *3/33*

## RESEARCH INTEREST

My current research interest is in designing algorithms and systems for managing and processing large-scale data efficiently and with low latency. In particular, I am interested in approximate query processing using random sampling, sketches and/or other techniques. I am also interested in building hybrid transaction and analytical processing (HTAP) database systems.

## AWARDS AND HONORS

1. Google PhD Fellowship 2019
2. Best Paper Award, ACM SIGMOD 2016
3. Zhiyuan Scholarship, Zhiyuan College, Shanghai Jiao Tong University 2016
   - Awarded to top 4 undergraduate students in the ACM Class
4. Kaiyuan Scholarship, Zhiyuan College, Shanghai Jiao Tong University 2014
5. First Prize, National Olympiad in Informatics in Province (NOIP) 2011

## PUBLICATIONS

*: Authors in alphabetic ordering

1. At-the-time and Back-in-time Persistent Sketches, by B. Shi, **Z. Zhao**, Y. Peng, F. Li, J. Phillips. In Proceedings of 40th ACM SIGMOD International Conference on Management of Data (**SIGMOD 2021**), pages TBD, Xi'an, Shaanxi, China, June 2021.
2. Efficient Join Synopsis Maintenance for Data Warehouse, by **Z. Zhao**, F. Li, Y. Liu. In Proceedings of 39th ACM SIGMOD International Conference on Management of Data (**SIGMOD 2020**), pages 2027-2042, Portland, USA, June 2020.
3. SolarDB: Towards a Shared-Everything Database on Distributed Log-Structured Storage, by T. Zhu, **Z. Zhao**, F. Li, W. Qian, A. Zhou, D. Xie, R. Stutsman, H. Li, H. Hu. In ACM Transactions on Storage (**TOS**), Volume 15, Issue 2, June 2019.
4. *Wander Join and XDB: Online Aggregation via Random Walks, by F. Li, B. Wu, K. Yi, **Z. Zhao**. In ACM Transactions on Database Systems (**TODS**), Volume 44, Issue 1, Article No. 2, January 2019.

5. InferSpark: Statistical Inference at Scale, by **Z. Zhao**, J. Pei, K. Zhu, E. Lo, C. Liu. In 2019 IEEE International Conference on Big Data and Smart Computing (**BigComp 2019**), pages 1-8, Kyoto, Japan, February 2019.
6. Solar: Towards a Shared-Everything Database on Distributed Log-Structured Storage, by T. Zhu, **Z. Zhao**, F. Li, W. Qian, A. Zhou, D. Xie, R. Stutsman, H. Li, H. Hu. In Proceedings of the 2018 USENIX Annual Technical Conference (**ATC 2018**), pages 795-807, Boston, USA, July 2018.
7. Random Sampling over Joins Revisited, by **Z. Zhao**, R. Christensen, F. Li, X. Hu, K. Yi. In Proceedings of 37th ACM SIGMOD International Conference on Management of Data (**SIGMOD 2018**), pages 1525-1539, Houston, USA, June 2018.
8. *Wander Join and XDB: Online Aggregation via Random Walks, by F. Li, B. Wu, K. Yi, **Z. Zhao**. In **SIGMOD Record**, Volume 44, Number 1, pages 33-40, May 2017.
9. *Wander Join: Online Aggregation via Random Walks, by F. Li, B. Wu, K. Yi, **Z. Zhao**. In Proceedings of 35th ACM SIGMOD International Conference on Managment of Data (**SIGMOD 2016, Best Paper Award**), pages 615-629, San Francisco, USA, June 2016.
10. *Wander Join: Online Aggregation for Joins, by F. Li, B. Wu, K. Yi, **Z. Zhao**. In Proceedings of 35th ACM SIGMOD International Conference on Management of Data, Demo Paper, (**SIGMOD 2016**), pages 2121-2124, San Francisco, USA, June 2016.

## PROFESSIONAL EXPERIENCE

**Data Group, University of Utah** Aug 2016 - Present
*Research Assistant*

· Advised by Prof. Feifei Li

· **Online Aggregation over Joins**

- Designed the Wander Join algorithm for generating non-uniform and independent samples of join results with replacement, which can be used for approximating SUM/AVG/COUNT aggregations
- Implemented Wander Join in various systems (in-memory, PostgreSQL, Apache Spark)
- Integrated the PostgreSQL version of Wander Join with Apache Zeppline to enable online aggregation over join queries

· **Random Sampling over Joins**

- Designed a new algorithmic framework for generating uniform and independent samples over joins
- Conducted extensive experiments against state-of-the-art algorithms
- Works for approximate query processing, down-sampling datasets for ML and etc.

· **Join sample maintenance in data warehouses and data streams**

- Designed algorithms and indexes for maintaining uniform and independent samples (w or w/o replacement) in streaming systems or updatable data warehouses
- Implemented a single-node database system for streaming join sampling and its maintenance
- Studying the design of indexes and concurrency control scheme for parallel and distributed join sample maintenance over data streams

· **At-the-time and back-in-time persistent sketches**

- Defined at-the-time (ATTP) and back-in-time (BITP) persistence for data sketches in contrast to those for arbitrary time windows
- Constructed ATTP and BITP sketches with higher accuracy and lower space cost
- Designed an experimental framework for evaluating ATTP and BITP sketches

**Google** May 2020 - Aug 2020
*Research Intern*

· SysInfra Performance team, supervisor: Dr. Yixin Luo
· Research on improving the performance of approximate query processing in F1

**Alibaba Group** May 2019 - Aug 2019
*Research Intern*

· Database department, supervisor: Dr. Sheng Wang
· Design and implement high availability components for PolarDB

**Microsoft Research** May 2018 - Aug 2018
*Resaerch Intern*

· DMX Group, Mentor: Dr. Srikanth Kandula
· Study the performance of random sampling on different storage media

**Microsoft Research** May 2017 - Aug 2017
*Research Intern*

· DMX group, Mentor: Dr. Chi Wang
· Research on random sampling in machine learning pipelines

**The Hong Kong Polytechnic University** August 2015 - January 2016
*Research Assistant*

· Advisor: Prof. Eric Lo
· Research on statistical inference on probabilistic graphical models with large-scale data

**Shanghai Jiao Tong University** August 2014 - May 2016
*Research Assistant*

· Advisor: Prof. Kenny Q. Zhu
· Research on probabilistic programming

## SELECT OPEN-SOURCE PROJECTS

- XDB (approXimate DataBase): Wander Join in PostgreSQL
- Random Sampling over Joins
- Efficient Join Synopsis Maintenance for updatable Data Warehouse

## TEACHING EXPERIENCE

- CS6530 Database Systems, University of Utah
  *Lecturer* Aug 2018 - Dec 2018
- CS6530 Database Systems, Univesrity of Utah
  *Teaching Assistant* Aug 2017 - Dec 2017
- Database Systems, Shanghai Jiao Tong University
  *Teaching Assistant* May 2016 - June 2016
- Introduction to Computer Science (Logic and Combinatorics), Shanghai Jiao Tong University
  *Teaching Assistant* Sept 2015 - Jan 2015
- Principle and Practice of Computer Algorithms, Shanghai Jiao Tong University
  *Teaching Assistant* July 2014