Jinyang Li

DOB: 02/28/1998　　　　**Email: lijinyang@pku.edu.cn**

EDUCATION BACKGROUND

Peking University　　　　　　　　　　　　　　　　　　　　Beijing, China

Major: Computer Science　　　　　　　　　　　　　　　　09/2016-06/2020

Degree: Bachelor of Science　　　　**GPA:** 3.619/4.000

Related Courses:

Mathematics: Linear Algebra, Advanced Mathematics, Set Theory and Graph Theory, Probability Theory and Statistics(A), Algebraic Architecture and Combinatorial Mathematics, Game Theory

Computer Science: Practice of Programming in C&C++, Data Structure and Algorithm(A), Introduction to Computer Systems, Algorithm Design and Analysis, Computer Architectures, Operating Systems(A), Introduction to Database Systems, Big Data Management Technology

Research Interests: databases, data analysis

Personal Homepage: https://lijinyang0228.github.io/

PROFESSIONAL EXPERIENCES

Project name: Data Spread　　　　　　　　　　　　　　07/2019 – Present

Individual Research, instructed by Prof. Kevin Chen-Chuan Chang, UIUC

Objective: Combine spreadsheets with databases to benefit from both

- ✧ Designed a data structure to maintain the mapping from positions of rows (to represent a row in spreadsheets) to rowIDs (to query a row in database), which supported queries, updates (insertions, deletions) and order manipulations (reorder, swap, move)
- ✧ Designed a self-balancing tree called *Stree* to deal with challenge 1: an update causes cascading updates of positions of other rows, and challenge 2: order manipulations need collective updates
- ✧ Designed *Dynamic Array* for challenge 3: adapt to unbalanced reads and writes in varying workloads
- ✧ These two data structures have achieved shorter average response time and more stable performance than traditional methods (standard arrays, linked list and Counted-Btree)

Outcome: I am preparing the paper draft and planning to submit to VLDB in December

Project: Windows Sketches　　　　　　　　　　　　　　01/2019 -- 05/2019

Instructed by Prof. Tong Yang, Peking University

Objective: Propose a data structure to estimate the frequency, persistency and recency of all items at the same time in a data stream

- ✧ Used counters to record persistency that emphasizes on the number of time windows, and logged the last time window an item appears to avoid duplicates, which also represents recency

Outcome: The final paper is in preparation (second author)

Project: Elastic Cuckoo filter　　　　　　　　　　　　　01/2019 -- 06/2019

Instructed by Prof. Tong Yang, Peking University

Objective: Overcome two shortcomings of the Cuckoo filter to get a full use of memory space

✧ Used a hash function to decide the range for each item, which made the size of a Cuckoo filter flexible without losing its elegance

✧ Extended a Cuckoo filter by copying and performing lazy updates: delete redundant copies in a bucket only when it is accessed in an insertion; shrank a Cuckoo filter by moving each fingerprint from bucket i to $i/2$

✧ Achieved the same performance as the standard Cuckoo filter but better space usage thanks to its elasticity

Outcome: The paper has been submitted to TPDS (first author)

PUBLICATIONS

- Fast and accurate stream processing by filtering the cold. Tong Yang, Jie Jiang, Yang Zhou, Long He, **Jinyang Li**, Bin Cui, Steve Uhlig, Xiaoming Li. **VLDBJ 2019**.
- HeavyKeeper: An Accurate Algorithm for Finding Top- k Elephant Flows. Tong Yang, Haowei Zhang, **Jinyang Li**, Junzhi Gong, Steve Uhlig, Xiaoming Li. **IEEE/ACM Transactions on Networking 2019**.

SKILLS

Programming language: C, C++, Java, Python, Go, Verilog, SQL

Computer skill: MapReduce, Hadoop, Hbase, Hive, Pig, ZooKeeper, NoSQL, Spark, GraphX