

Hao Jiang, PhD

Postdoctoral Fellow

Harvard John A. Paulson

School Of Engineering And Applied Sciences

150 Western Ave, Boston, MA 02134

(315)267-6271

hajang@seas.harvard.edu

<http://people.cs.uchicago.edu/~hajang>

Education

- **The University of Chicago** Chicago, IL
Ph.D. Computer Science 2015 – 2021
– Advisor: Aaron J. Elmore
- **Clarkson University** Potsdam, NY
M.Sc. Computer Science 2012 – 2015
– Advisors: Jeanna N. Matthews
- **Fudan University** Shanghai, China
B.Sc. Computer Science 2001 – 2005
– Advisor: Liang Zhang

Publications

1. Chunwei Liu, **Hao Jiang**, John Paparrizos, Aaron J. Elmore *Decomposed Bounded Floats for Fast Compression and Queries VLDB 2021*
2. **Hao Jiang**, Chunwei Liu, John Paparrizos, Aaron J. Elmore *Good to the last bit: Data-Driven Encoding with CodecDB SIGMOD 2021*
3. **Hao Jiang**, Chunwei Liu, Qi Jin, John Paparrizos, Aaron J. Elmore, *PIDS: Attribute Decomposition for Improved Compression and Query Performance in Columnar Storage, VLDB 2020*
4. Chunwei Liu, McKade Umbenhowe, **Hao Jiang**, Aaron J. Elmore, *Mostly Order Preserving Dictionaries, ICDE 2019*
5. **Hao Jiang**, Aaron J. Elmore, *Boosting Data Filtering on Columnar Encoding with SIMD, DaMon 2018*
6. Dixin Tang, **Hao Jiang**, Aaron J. Elmore, *Adaptive Concurrency Control: Despite the Looking Glass, One Concurrency Control Does Not Fit All, CIDR 2017*
7. **Hao Jiang**, Yaoqing Liu, Jeanna N. Matthews, *IP Geolocation Estimation using Neural Networks with Stable Landmarks, IEEE INFOCOM Workshop GI 2016*
8. Wenjin Hu, Long Zhang, **Hao Jiang**, Jeanna N. Matthews, *A Quantitative Study of Virtual Machine Live Migration, CAC 2013*

Research Experience

- Harvard University**

Cambridge, MA

Postdoctoral Researcher with Prof. Stratos Idreos

Oct. 2021 – Present

 - **Database in Blockchain** Explore the importance of the state datastore in a blockchain system and its impact to the performance
- The University of Chicago**

Chicago, IL

Research Assistant under Prof. Aaron J. Elmore

Sept. 2015 – Aug. 2021

 - **CoLSM: Lightweight Encoding in LSM-trees** Explore using lightweight encodings to LSM-trees to improve lookup efficiency. Design a LSM-tree that supports using different data structure at each level to improve merge performance.
 - **PIDS: Exploring Unsupervised Pattern Inference in String Attributes** We develop a new compression algorithm that discover patterns in string attributes, and uses the pattern to extract and compress sub-attributes independently.
 - **Encoding-Aware Columnar Database** Design and build a columnar database that uses encoding knowledge to speed up database queries.
 - **Data-Driven Database Encoding Selection.** Using Machine Learning Techniques to select most efficient data encoding schema for a given columnar database.
 - **SBoost: Using SIMD Instruction to speed up Database Operator.** Explore the possibility of speed up database joining and scanning operation using SIMD Instructions.
 - **Stream Partitioning on Large-Scale Graph.** Design a new algorithms for Stream Partitioning of Large-Scale Graph, which has been proven to outperform current state-of-art algorithm while maintaining time efficiency.
 - **Pattern as a Foreign Language.** Using Recurrent Neural Network to discover and extract hidden pattern from database columns. **Graph Classification** Design a sampling based method to classify whether a Large-Scale Graph satisfy power-law distribution. This result is further used to build an adaptive partitioning method that can distinguish power-law and non-power-law graphs and apply different partitioning method.
 - **Distributed Storage on ZCCloud.** Design and implement a simulator that study the system availability and performance of Cassandra distributed key-value store running on Zero-Cabon Cloud Datacenter platforma using stranded power.
- Clarkson University**

Potsdam, NY

Research Assistant under Prof. Jeanna N. Matthews

Sept. 2012 – Jul. 2015

 - **Internet Toplogy Platform.** Design, implement and experiment several heuristic based partitioning methods on Internet Topology. Experiment on building a Internet Topology Platform support thirdparty data analysis programs to access Internet structure with ease.
 - **GreenDataCenter Project.** Participate in GreenDataCenter (GDC) Project. Design and implement a simulation environment to study the feasibility of using pure green energy, such as solar and wind to power distribute data center and provide service with high availability.
 - **IP Geolocation.** Proposed an innovative method of locating the physical location of arbitrary IP address with a two-tier neural network. Independently designed and implemented the system.
 - **Phishing website detection with logo recognition.**Used SVM to extract the embedded logos of known websites from a large image to identify phishing websites based on these logos. Independently designed and implemented the system.
- Fudan University**

Shanghai, China

Research Assistant under Prof. Liang Zhang

Sept. 2003 – Jul. 2005

 - **Grid Computing.** Study the feasibility of using IBM Globus Platform to build a grid-computing system for bio-informatic computation and database service.

Work Experience

- **Harvard University** Cambridge, MA
Postdoctoral Fellow Oct. 2021 – Present
 - Research on Permissioned Blockchain performance
- **Facebook Inc.** Menlo Park, CA
Software Engineer Intern July. 2018 – Sept. 2018
 - Design, implement and evaluate a load balance algorithm for PHP requests
- **Baidu Inc.** Shanghai, China
System Architect Sept. 2011 – Aug. 2012
 - Design and implement a customizable MySQL replication framework. This framework intercepts message sent by a master database during MySQL master-slave replication, rewrite it with user provided function and send it to slave database. This framework allows an easy customization of replication behaviors.
 - BigData Analysis System using Hadoop. Lead a 4-developer team to design and develop a Hadoop-based BigData analysis system. The system processes a daily data volume of over 10 terabytes.
- **OOCL Co. Ltd.** Shanghai, China
Senior Software Engineer Sept. 2005 – Sept. 2011
 - Production Server JVM Performance Tuning. Design and implement a log analysis system for production environment JVM resource leak and memory leak tracing and tuning.
 - Design and optimization of an accounting system containing billions of rows and processing millions of queries hourly based on Oracle DBMS.

Teaching Experience

- **Introduction to Database** The University of Chicago
Teaching Assistant Spring 2018
 - Holding office hours and answer student questions
- **Computer Security** The University of Chicago
Teaching Assistant Autumn 2015, Spring 2016
 - Holding office hours and answer student questions
- **Introduction to Computer Science** Clarkson University
Teaching Assistant Fall, 2014
 - Holding Labs, lectures and recitation classes
- **Calculus I** Clarkson University
Teaching Assistant Fall 2013
 - Led recitation classes guiding students through difficult problems and reviewing course material.
- **Linear Algebra and Differential Equations** Clarkson University
Teaching Assistant Fall 2012
 - Proctoring exams and grading

Presentations

- *CodecDB: A Compression Aware Columnar Database*
 - SIGMOD 2021, Online, June 2021.
- *Attribute Decomposition for String Compression*
 - VLDB 2020, Online, June 2020.
- *Efficient Query on Compressed Data with SIMD*
 - DaMoN 2018, Austin, TX. June 2018.
- *Data-Driven Lightweight Encoding Selection*
 - CERES Research Summit - The University of Chicago, Chicago, IL. March 2017.
- *An Log-based Dynamic Partitioning Method*
 - System Seminar - The University of Chicago CS System Group, Chicago, IL. Oct 2016.
- *IP Geolocation using Two-tier Neural Network*
 - 6th Global Internet Symposium, San Francisco, CA. Apr 2016.

Participation in Workshops and Conferences

- | | |
|---|---------------------------------|
| • ACM SIGMOD
<i>SIGMOD 2021</i> | Online
June 2021 |
| • Very Large Database
<i>VLDB 2020</i> | Online
June 2020 |
| • Very Large Database
<i>VLDB 2019</i> | Los Angeles, CA
May 2019 |
| • ACM SIGMOD
<i>SIGMOD/PODS 2018</i> | Houston, TX
May 2018 |
| • ACM SIGMOD/PODS
<i>SIGMOD/PODS 2017</i> | Chicago, IL
May 2017 |
| • 1st Workshop on Data Management and End-to-End Machine Learning
<i>DEEM 2017</i> | Chicago, IL
May 2017 |
| • Center for Unstoppable Computing Research Summit
<i>The University of Chicago</i> | Chicago, IL
March 2017 |
| • Quantum Computing Symposium
<i>Argonne National Lab</i> | Argonne, IL
May 2016 |
| • 6th Global Internet Symposium
<i>IEEE INFOCOM 2016</i> | San Francisco, CA
April 2016 |

Technical Skills

- Computer Science
 - Machine Learning, Deep Learning, Optimization, Parallel Programming, Database, Distributed Ssytems, Data Structure
- Programming Languages
 - Java, Scala, C/C++, Python, Matlab, Javascript, HTML/CSS