Hao Jiang
The University of Chicago
1100 E. 58th St. Chicago,IL, 60637

(315)267-6271
hajiang@cs.uchicago.edu
http://people.cs.uchicago.edu/~hajiang

Introduction

I am a sixth year Ph.D. student in Compute Science at The University of Chicago. I work primarily on Database, Distributed Systems and Machine Learning. My current research direction is exploring compression techniques in database.

Publications

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

Education

- **The University of Chicago** — Chicago, IL
  Ph.D. Computer Science (in progress) — *2015 – present*
  - Advisor: Aaron J. Elmore
  - Research Direction: Database & Machine Learning
- **Clarkson University** — Potsdam, NY
  M.Sc. Computer Science — *2012 – 2015*
  - Advisors: Jeanna N. Matthews
  - Research Direction: Distributed System
  - Graduated GPA 4.0
- **Fudan University** — Shanghai, China
  B.Sc. Computer Science — *2001 – 2005*
  - Advisor: Liang Zhang

– Major GPA 3.75

Research and Work Experience

- **Facebook Inc.** Menlo Park, CA
  Software Engineer Intern *July. 2018 – Sept. 2018*
  – Design, implement and evaluate a load balance algorithm for PHP requests
- **The University of Chicago** Chicago, IL
  Research Assistant under Prof. Aaron J. Elmore *Sept. 2015 – Present*
  – PIDS: Unsupervised Pattern Inference and Compression for String Attributes
  – Use SIMD to speed up scanning on bit-packed data
  – Use Neural Network to optimize Lightweight Encoding for Columnar Database.
  – Design a new algorithms for Stream Partitioning of Large-Scale Graph
  – Design a sampling based method to classify whether a Large-Scale Graph satisfy power-law distribution
- **Clarkson University** Potsdam, NY
  Research Assistant under Prof. Jeanna N. Matthews *Sept. 2012 – Jul. 2015*
  – 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.
  – 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.
- **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.

Skills

- **Computer Science:** Machine Learning, Optimization, Parallel Programming, Database, Distributed Ssytems, Data Structure
- **Tech Stack:** J2EE, Java, Scala, C/C++, Python, Matlab, Javascript, HTML/CSS
- **Teamworking:** Experience of leading small development teams