Hongzhi CHEN

Zhonghang Square, Beijing, China, ByteDance Inc.
(+86) 150-027-34771
chzyaobaiwei@gmail.com
https://yaobaiwei.github.io/

Education

15.8 - 20.7 **The Chinese University of Hong Kong (CUHK)** *Shatin,* Hong Kong
Ph.D., Computer Science and Engineering
- Supervisor: Prof. James Cheng

11.9 - 15.6 **Huazhong University of Science and Technology (HUST)** *Wuhan,* China
B.Eng, Computer Science and Technology
Overall GPA: 3.88 / 4.00 Rank: 1/ 30/ 472 (**in a 30-student Honor Class**)

Work Experiences

20.7 - present **ByteDance Inc, *Senior Researcher and System Architect, ByteDance Infra*** Beijing, China
- Idea Proposer, Tech Lead, Contributor of **ByteGNN**: a high-performance end-to-end GNN training system at scale, performs at most 30+X speedup compared with the open-source SOTA (e.g., Euler, GraphLearn, DGL).
- Idea Proposer, Tech Lead, Contributor of **ByteGAP**: an out-of-core graph analysis system with DRAM/NVM hybrid store design.
- Worked on the system designs and optimizations of **ByteGraph**, a distributed graph database which serves 500+ online business at Bytedance.
- Leading all research works about graph training/processing/storage at ByteDance.

19.2 - 19.8 **Huawei, 2012 Lab,** *Parallel and Distributed Computing Laboratory* Shenzhen, China
- **Researcher**, System Design for **MindSpore**.
  Worked on the prototype of MindSpore, a self-developed distributed deep learning system at 2012-Lab@Huawei. It has come the most impactful DL/AI platform and community in China.

Academic Experiences

17.05 - 17.08 **University of Pennsylvania,** *NetDB Lab, Dept. CIS* P.A., U.S.A.
- **Visiting Scholar**, Distributed Graph Query Optimizations over Data Centers.
  Supervisor: **Prof. Boon Thau Loo**.

15.09 - 20.07 **The Chinese University of Hong Kong,** *HDL Lab, Dept. CSE* Hong Kong
- **Research Assistant**, Distributed Systems, Distributed Database, Graph Systems.
  Supervisor: **Prof. James Cheng**.

14.06 - 15.05 **Microsoft Research Asia,** *Software Analytics Group* Beijing, China
- **Research Intern**, Large-scale Data Analytics over Cosmos, Distributed Data Mining.
  Supervisor: **Qingwei Lin** (Lead Researcher) and **Dr. Jianguang Lou** (Principle Researcher).

13.09 - 14.06 **HUST,** *IDC Lab, Dept. CSE* Wuhan, China
- **Research Intern**, System Optimization for Hadoop's Map-Reduce.
  Supervisor: **Prof. Ruixuan Li**

Research Interests and Projects

My general research interests cover the broad area of distributed system and database, with special emphasis on distributed graph training/processing/storage systems and distributed machine learning/deep learning systems. My current works focus on building solid and practicable graph data warehouse and graph AI system in industry.

Many of my past projects have been deployed at ByteDance and Microsoft or inspired the system designs in industrial companies. I led or participated the following projects:

ByteGNN	An end-to-end scalable GNN traning system with interactive functionality based on Ray.
ByteGAP	An out-of-core graph analysis system with DRAM/NVM hybrid store design to support trillion graph.
ByteGraph	A high-performance distributed graph database, which supports 500+ OLTP workloads at Bytedance.
G-Tran	A high-performance MVCC-based in-memory graph database with strong isolation support.
Grasper	An RDMA-enabled OLAP system over property graphs with good scalability.
G-Miner	A distributed graph mining system aimed at general graph mining problems.
PPA-assembly	A scalable toolkit for de novo genome assembly was developed based on Pregel.
GraphD	It offers out-of-core support for processing very big graphs in a small cluster of commodity PCs.
LWCP	A fault tolerance mechanism for Pregel-like systems with performance tens of times faster than the baselines.
GraphRex	An efficient framework for graph processing on datacenter infrastructure.
Pregel+	A Pregel-like system with optimizations to reduce communication cost and eliminate skewness in communication.

Service-Intelligence: A distributed log mining tool based on Microsoft Cosmos, aiming to do Text Clustering and Anomaly Detection based on large streaming data over Azure, paper published on **ICSE'16**.

Service-Insider: A data mining algorithm package in EXCEL for Frequent Pattern Mining, Event Clustering, Association Rule Mining, Anomaly Detection, Mutil-Dimension Change Detection.

iDice: An efficient algorithm for Emerging Issues Finding on mutli-dimension data, paper published on **ICSE'16**.

In4: A distributed OLAP system based on Actor model to support online data analysis and mining, paper published on **SIGKDD'18**.

I am fortunate to mentor 10+ bright Ph.D./M.Phil./M.SC./UG interns from CUHK, UPenn, THU, PKU, HUST, HIT, BIT, etc.

Publications

[1] *G-Tran: A High Performance Distributed Graph Database with a Decentralized Architecture* **VLDB'22**
Hongzhi Chen, Changji Li, Chenguang Zheng, Chenghuan Huang, Juncheng Fang, James Cheng, Jian Zhang.

[2] *ByteGraph : A High-Performance Distributed Graph Database in ByteDance* **VLDB'22**
Changji Li, **Hongzhi Chen***, Shuai Zhang, Yingqian Hu, Chao Chen, Zhenjie Zhang, Meng Li, Xiangcheng Li, Dongqing Han, Xiaohui Chen, Xudong Wang, Huiming Zhu, Xuwei Fu, Tingwei Wu, Hongfei Tan, Hengtian Ding, Mengjing Liu, Kangcheng Wang, Ting Ye, Lei Li, Xin Li, Yu Wang, Chenguang Zheng, Hao Yang, James Cheng.

[3] *Colorful h-star Core Decomposition* **ICDE'22**
Sen Gao, Ronghua Li, Hongchao Qin, **Hongzhi Chen**, Ye Yuan, Guoren Wang.

[4] *Fast Maximal Clique Enumeration on Uncertain Graphs: A Pivot-based Approach* **SIGMOD'22**
Qiangqiang Dai, Ronghua Li, Meihao Liao, **Hongzhi Chen**, Guoren Wang.

[5] *Lightning Fast and Space Efficient k-clique Counting* **WWW'22**
Xiaowei Ye, Ronghua Li, Qiangqiang Dai, **Hongzhi Chen**, Guoren Wang.

[6] *ByteGNN: Efficient Graph Neural Network Training at Large Scale* **VLDB'22**
Chenguang Zheng, **Hongzhi Chen***, Yuxuan Cheng, Zhezheng Song, Yifan Wu, Changji Li, James Cheng, Hao Yang, Shuai Zhang.

[7] *BGL: GPU-Efficient GNN Training by Optimizing Graph Data I/O and Preprocessing* **arXiv'21**
Tianfeng Liu, Yangrui Chen, Dan Li, Chuan Wu, Yibo Zhu, Jun He, Yanghua Peng, Hongzheng Chen, **Hongzhi Chen**, Chuanxiong Guo.

[8] *G-Tran: Making Distributed Transactions over Big Graphs Fast* **arXiv'21**
Hongzhi Chen, Changji Li, Chenguang Zheng, Chenghuan Huang, Juncheng Fang, James Cheng, Jian Zhang.

[9] *High Performance Distributed OLAP on Property Graphs with Grasper* **SIGMOD'20**
Hongzhi Chen, Bowen Wu, Shiyuan Deng, Chenghuan Huang, Changji Li, Yichao Li, James Cheng.

[10] *Measuring and Improving the Use of Graph Information in Graph Neural Networks* **ICLR'20**
Yifan Hou, Jian Zhang, James Cheng, Kaili Ma, Richard T. B. Ma, **Hongzhi Chen**, Ming-Chang Yang.

[11] *Grasper: A High Performance Distributed System for OLAP on Property Graphs* **SoCC'19**
Hongzhi Chen, Changji Li, Juncheng Fang, Chenghuan Huang, James Cheng, Jian Zhang, Yifan Hou, Xiao Yan.

[12] *A Representation Learning Framework for Property Graphs* **SIGKDD'19**
Yifan Hou, **Hongzhi Chen**, Changji Li, James Cheng, Ming-Chang Yang.

[13] *Large Scale Graph Mining with G-Miner* **SIGMOD'19**
Hongzhi Chen, Xiaoxi Wang, Chenghuan Huang, Juncheng Fang, Yifan Hou, Changji Li, James Cheng.

[14] *Optimizing Declarative Graph Queries at Large Scale* **SIGMOD'19**
Qizhen Zhang, Akash Acharya, **Hongzhi Chen**, Simran Arora, Ang Chen, Vincent Liu, Boon Loo.

[15] *Scalable De Novo Genome Assembly Using a Pregel-Like Graph-Parallel System* **TCBB'19**
Guimu Guo, **Hongzhi Chen**, Da Yan, James Cheng, Jake Chen, Zechen Chong.

[16] *Lightweight Fault Tolerance in Pregel-Like Systems* **ICPP'19**
Da Yan, James Cheng, **Hongzhi Chen**, Cheng Long, Purushotham Bangalore.

[17] *G-Miner: An Efficient Task-Oriented Graph Mining System.* **EuroSys'18**
Hongzhi Chen, Miao Liu, Yunjian Zhao, Xiao Yan, Da Yan, James Cheng.

[18] *Scalable De Novo Genome Assembly Using Pregel.* **ICDE'18**
Da Yan, **Hongzhi Chen**, James Cheng, Zhenkun Cai, Bin Shao.

[19] *GraphD: Distributed Vertex-Centric Graph Processing Beyond the Memory Limit.* **TPDS'18**
Da Yan, Yuzhen Huang, Miao Liu, **Hongzhi Chen**, James Cheng, Huanhuan Wu, Chengcui Zhang.

[20] *Norm-Ranging LSH for Maximum Inner Product Search.* **NIPS'18**
Xiao Yan, Jinfeng Li, Xinyan Da, **Hongzhi Chen**, and James Cheng.

[21] *Architectural Implications on the Performance and Cost of Graph Analytics Systems.* **SoCC'17**
Qizhen Zhang, **Hongzhi Chen**, Da Yan, James Cheng, Boon Thau Loo, Purushotham Bangalore.

[22] *G-thinker: Big Graph Mining Made Easier and Faster.* **arXiv'17**
Da Yan, **Hongzhi Chen**, James Cheng, M.Tamer.Ozsu, Qizhen Zhang, John C.S. Lui.

Awards & Honors

Fall 2019- Spring 2020 **Researcher@Microsoft Research Asia**, **TopMinds@Huawei**, **AliStar@Alibaba-Damo**, **TechMaster@Tencent**.
2019.11 SoCC'19 Travel Award
2019.5 SIGMOD'19 Travel Award
2018.4 EuroSys'18 Travel Award
2016 - 2020 CUHK Postgraduate Studentship.
2015.6 Award Winner of **Hong Kong PhD Fellowship**.
2015.6 **"Stars of Tomorrow" at Microsoft Research Asia** (Only 15% research interns won the Award)
2015.6 Outstanding Graduates (3% in HUST)
2014.10 **Top-100 Excellent Undergraduates in Computer Science** by China Computer Federation (**Top 0.1%**)
2014.9 Academic Excellence Scholarship (2% in HUST)
2013.9 National Undergraduate Scholarship (2% in HUST)
2012.9 Academic Excellence Scholarship (2% in HUST)
2012.9 Most Outstanding Undergraduate (1% in HUST)

Teaching

Spring, 2020 CSCI5120: Advanced Topics in Database Systems
Spring, 2018 CSCI1020: Hands-on Introduction to C++
Fall, 2017 ENGG1110: Problem Solving By Programming
Spring, 2017 ENGG1110: Problem Solving By Programming
Fall, 2016 ENGG1110: Problem Solving By Programming

Professional Activities

External Reviewer

2020 SIGMOD
2019 SIGMOD
2018 VLDB, ICDE
2017 VLDB, ICDE, CCGRID, BigData
2016 VLDB, KDD, SOCC, ICDM, DASFAA, BigData, APWeb

Participation and Talks

2019 ACM Symposium on Cloud Computing, Santa Cruz, California, U.S.A.
2019 International Conference on Management of Data, Amsterdam, Netherlands
2018 European Conference on Computer Systems, Porto, Portugal
2015 China National Computer Congress, Zhengzhou, China