Hongzhi CHEN

Zhonghang Square, Beijing, China, ByteDance Inc.
✆ *(+86) 150-027-34771*
✉ *chenhongzhi@bytedance.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**)

Research Interests

My general research interests cover the broad area of distributed systems and databases, with special emphasis on distributed graph systems and distributed machine learning/deep learning systems. My current works focus on RDMA based OLAP/OLTP systems over graphs, graph databases, distributed Graph Neural Network systems.

Work Experiences

20.7 - present	**ByteDance Inc,** *Senior R.D.@Infrastructure Team*	Beijing, China

○ Leading a team to achieve high-performance large-scale GNN training without GPU, currently with at most 30+X speedup compared with the open-source SOTA (e.g., Euler, GraphLearn, DGL).
○ Working on system design and implmentation of **ByteGraph**, which is ByteDance's company-level Distributed Graph Database where holding 400+ inner scenarios.
○ Taking in charge of all research works about Graph Processing/Training/Storage.

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. Now, 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**

Publications

[1] *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.

[2] *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.

[3] *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.

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

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

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

[7] *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.

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

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

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

[11] *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.

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

[13] *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.

[14] *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.

Projects during my PhD and Undergraduate Study

My research focuses on the design and implementation of distributed systems and their applications. I have led or participated as the core developer of the following projects.

G-Tran	A high-performance MVCC-based graph database for processing transaction over large-scale graph with strong isolation support.
Grasper	An RDMA-enabled high performance OLAP system over property graphs with good scalability. We proposed a novel execution engine, called Expert Model, to provide tailored optimizations with adaptive parallelism control for query operators. Grasper achieves order of magnitudes performance improvements over existing systems (e.g., Titan, JanusGraph, OrientDB, Neo4j).
PGE	A representation learning framework for property graph embedding. The key idea of PGE is a three-step framework to leverage both the topology and property information into Graph Neural Networks for a better node embedding result.
G-Miner	A distributed graph processing system aimed at general graph mining problems, which have intensive local computation inside a subgraph. We modeled each subgraph processing as a task and designed a task-based pipeline to improve the parallelism between computation and communication. A dynamic task stealing mechanism as well as an efficient cache strategy were also proposed to further speed up the task processing.
G-thinker	Real applications, such as graph matching and community detection, often require computation intensive graph analytics, which cannot be represented by vertex-centric algorithms for efficient execution in systems like Pregel and GraphLab. We proposed G-thinker, a new subgraph-centric general graph processing distributed system, which is natural for subgraph finding problem.
PPA-assembly	A scalable toolkit for de novo genome assembly was developed based on Pregel. PPA-Assembly provides a set of key operations in genome assembly, which were implemented by practical Pregel algorithms (PPAs) with strong performance guarantees.
GraphD	It offers out-of-core support for processing very big graphs in a small cluster of commodity PCs, with performance comparable with the state-of-the-art distributed in-memory graph systems.
LWCP	A fault tolerance mechanism for Pregel-like systems with performance tens of times faster than the conventional checkpointing mechanisms.
FPMS	A novel and general distributed framework to mine frequent patterns, including frequent item-sets/sequences/graphs.

I also joined in some other research projects during my work at CUHK, UPenn and MSRA.

GraphRex	An efficient framework for graph processing on datacenter infrastructure. The key technical contribution of GraphRex is the identification and optimization of a set of global operators whose efficient implementation is crucial to the good performance of large, datacenter-based graph analysis.
RANGE-LSH	Using maximum inner product for similarity search, which significantly outperforms SIMPLE-LSH. RANGE-LSH is robust to the shape of 2-norm distribution and different partitioning methods.
Pregel+	An open-source Pregel implementation with optimizations to reduce communication cost and eliminate skewness in communication.

More If you are interested in my previous works at **Microsoft Research**, **Software Analytics Group**, please visit my homepage for more details. My works at MSRA focused more on distributed data analytics, including OLAP, pattern mining, text clustering and anomaly detection. I participated in 4 projects as a core developer, i.e. *Service-Intelligence*, *Service-Insider*, *iDice* and *In4*. In particular, both **Service-Intelligence** and **iDice** have been published in **ICSE'16**, and **In4** has been published in **KDD'18**.

Awards & Honors

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	The original 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	CCF (China Computer Federation) National **Top 100** Outstanding Undergraduates (**Top 0.1%**)
2014.9	Academic Excellence Scholarship (2% in HUST)
2014.9	Merit Undergraduate (2% in HUST)
2013.9	National Undergraduate Scholarship (2% in HUST)
2013.9	Merit Undergraduate (2% in HUST)
2012.9	Most Outstanding Undergraduate (1% in HUST)
2012.9	Academic Excellence Scholarship (2% 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