

Zeyuan Hu

✉ zeyuan.zack.hu@gmail.com 🏠 <https://zhu45.org> 🌐 <https://github.com/xxks-kkk> ☎ 512-200-5892

EDUCATION

- | | | |
|--|--------------------|-----------------------------|
| University of Texas | Austin, TX | Sept 2020 – May 2025 |
| <ul style="list-style-type: none">• Ph.D. in Computer Science.• Specialization: database (relational & graph) query optimization and processing | | |
| University of Texas | Austin, TX | Sept 2017 – May 2019 |
| <ul style="list-style-type: none">• M.S. in Computer Science. (GPA: 3.81/4.00) | | |
| University of Wisconsin | Madison, WI | Sept 2010 – Dec 2014 |
| <ul style="list-style-type: none">• B.A. Triple Major, Computer Science, Economics with Honors, Mathematics. (Major GPA Avg: 3.80/4.00) | | |

WORK EXPERIENCE

- | | | |
|---|---------------------------------------|-----------------------------------|
| Software Dev Engineer II
AWS Timestream Query Team | Amazon Web Services | January 2020 – August 2020 |
| <ul style="list-style-type: none">• Optimized Hot Tier (HT) query evaluation strategy so that <u>Presto</u> performed intelligent query pushdown and reduced query execution time by 22~29%• Modified Presto parser and optimizer to ensure only selected data types and SQL functions are picked and exposed to the customer• Added end-to-end <u>INTEGER</u> data type support• Combined Timestream internal catalog service with Presto to implement <u>SHOW DATABASES</u> and <u>SHOW TABLES</u> metadata queries• Redesigned the semantics of date & time support in Timestream such that customers can conceptually gauge the correctness of query with date & time string or timestamp | | |
| Cloud Architect Engineer
Omnia Storage Team | State Street Financial Service | June 2019 – November 2019 |
| <ul style="list-style-type: none">• Built auto-deployment system of IBM Cloud Object Storage in multi-site clusters using <u>Ansible</u> and <u>Docker</u>• Developed a distributed workload generator and performance benchmark toolkit written in <u>Go</u> with <u>Redis</u>, <u>InfluxDB</u>, <u>MongoDB</u>, and <u>Elastic Search</u> | | |
| Software Engineer
DB2 LUW Federation Team | IBM | August 2015 – August 2017 |
| <ul style="list-style-type: none">• Constructed <u>Hive</u> and <u>Impala</u> wrappers with <u>C++</u> and <u>Java</u> to support federation database between traditional RDBMS and Hadoop-based data warehouse solution• Created automated setup tools with <u>Shell</u> that reduce product configuration time by 75%• Enhanced server option optimization tools using <u>C</u> to reduce federation database performance tuning time by 90 % and enable the capability of tuning the product against Hive, Impala, and Spark• Resolved over 20 defects, including a severe memory leak issue that impacted a \$1.6 million deal | | |
| Software Engineer Internship
HPC Infrastructure Team | Schlumberger | May 2018 – August 2018 |
| <ul style="list-style-type: none">• Implemented a monitoring component of the in-house High-Performance Computing (HPC) engine in <u>C++</u> to provide the fault tolerance and handle the “straggler” problem• Employed SGD algorithm to dynamically learn the best timing for backup executions of the in-progress tasks based on the computation task characteristics• Built a <u>C++</u> code generator that automatically generates the application layer code based on the engine API | | |

- Teaching Assistant in Database Systems (Spring 2019), Data Engineering (Fall 2018, Fall 2020), Quantitative Methods in Neuroscience (Spring 2018), Differential Calculus (Fall 2017)

HONORS AND AWARDS

- 2017 IBM Appreciation program for the Practice: Dare to create original ideas, IBM
- 2016 IBM Manager's Choice Award - Put the Client First, IBM
- 2018 Best Internship Project Award (Software Engineering), Schlumberger
- 2014 Graduation with Distinction, University of Wisconsin
- 2013 Honors Summer Sophomore Research Apprenticeship, University of Wisconsin
- 2012 Meek Bishop Scholarship in Economics, University of Wisconsin
- 2010-2012 Dean's List, University of Wisconsin

SERVICE AND SOCIETIES

- IBM Diamond & Ring Toastmaster Club (Jun 2016 - Jun 2017), President
- UTCS Master Admission Committee (Jan 2018 – March 2018), Member

SPECIALIZED SKILLS

- **Languages:** C++, C, Python, Go, Rust, Shell, SQL, Java, Lisp, MATLAB
- **Software:** CMake, Autotools, QEMU, Docker, Tensorflow, Keras, Git, ClearCase, Hive, Impala, Hadoop
- **Graduate Coursework:** Machine Learning, Structured Models for NLP, Human Computation & Crowdsourcing, Natural Language Processing, Semantics, Distributed Systems, Advanced Operating Systems, Data Centers, Algorithms, Automated Logic Reasoning,

PAPERS

- Jialin Wu, **Zeyuan Hu**, Raymond J. Mooney. "Jointly Generating Captions to Aid Visual Question Answering". (ACL 2019 *Oral*)
- Jialin Wu, **Zeyuan Hu**, Raymond J. Mooney. "Joint Image Captioning and Question Answering" In *VQA Challenge and Visual Dialog Workshop at the 31st IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2018)

SELECTED PROJECTS & RESEARCH EXPERIENCE

- **RustFS** (2018 - 2019). Building a user-space file system that leverages NVMe SSD. [Rust](#), [SPDK](#)
- **Strata with Lease** (2018). Extended Strata file system with Lease mechanism to support concurrent file access across processes. [C](#)
- **HyperPebblesDB** (2018). Constructed a key-value store that is part of LevelDB family with focus on reducing write amplification. [C++](#), [CMake](#), [Autotools](#)
- **Distributed Key-Value Store** (2018). Built a distributed key-value store with [Python](#) that uses *eventually consistency* model with two session guarantees: *Read Your Writes* and *Monotonic Reads*.
- **SGX Benchmark** (2018). Used a cloud service benchmark (*CloudSuite*) to measure the performance penalty brought by SGX on the IPC between a web server (Nginx) and a PHP application in a local environment
- **Benchmark Journaling Write Amplification** (2018). Measured journaling impact on the write amplification of various file systems using customized [filebench](#), [blktrace](#), [iostat](#), and [strace](#)
- **Benchmark Write Amplification on Fragmented File Systems** (2018). Measured the write amplification of workloads from *filebench* and Git workload from *BetrFS* on fragmented (i.e., *aged*) file systems, which are created from file system aging tool (e.g., *Geriatric*)