

With all my formative work and research experiences, I found my genuine interest lay in operating systems related, covering distributed systems, high-performance computing, and storage device research. Motivated and dedicated, I look forward to exploring the breadth and width of the computer systems fields through systematic postgraduate training. After completing the Graduate degree, I plan to become a computer scientist at a research institution or university, innovating high-performance system platforms on enormous data scales. Prospectively, I wish to materialize my vision of accelerating data responsiveness and computational speed at the system level, benefiting all data-oriented work and the information industries.

My interest in computer systems originated from a massive distributed system crash during my second internship. When Baidu acquired YY, the network platform I worked on collapsed due to the internally segmented cloud services and unprocessed big data. At that moment, I realized how data analysts' work was significantly correlated with and influenced by distributed systems. Regardless of individual capability, all the data analysts can do during a system pause is wait for system engineers to optimize and fix the system at the lower level. After proactively learning the huge impact and improvement of computer storage and computing on system performance, my professional pursuit was directed to a computer systems engineer and scientist, where I could understand and transform the basis of all data in greater depth.

Guided by my passion and ambition towards the computer system field, I get the privilege of working with **Prof. Yang Wang**, my operating systems course instructor at **The Ohio State University**, and **Prof. Michael D. Bond** on the **Online-ZebraConf** project. This is my first system research project aimed at developing a framework to modify existing unit tests to find heterogeneous unsafe configuration parameters in real-world applications. **Online-ZebraConf** solves the problem of not having specific control over reconfiguring parameters on multiple nodes at the same time in the existing solution. By introducing synchronization mechanisms like Timed modification within different nodes, Online-ZebraConf eliminates the racing effect between reconfiguration requests and synchronization correctness. In addition, I developed an automatic online testing pipeline that reconfigured the Hadoop cluster without manual interference and found more than 30 parameters related to synchronization correctness within the datanode. Automatic parameter inference is challenging due to the significant number of configuration parameters. While modifying the distributed system, I got exposed to the automated testing procedure constructions and potential system performance improvement approaches, such as network and storage.

My systems research continues at **the University of Chicago**, where I worked with **Prof. Haryadi S. Gunawi** on exploring tail latency reduction in storage devices. Given my background in data analysis and my interest in computer systems, I really enjoyed the project. I work on a flash latency cutting project called Flashnet, aiming to build an open ML training ground for I/O latency prediction. Inspired by ImageNet, it is planned to decentralize the data collection and data processing part to the storage and ML personnel. Building **Flashnet** is challenging because I need to understand both **storage and ML** domains. During my work on **Flashnet**, I recognized a very important problem: the data is unbalanced, as most data lie in the low-latency range and

very few outliers data are in the high-latency range. I accordingly pointed out the importance of the false negative rate metric because compared to accuracy, correctly identifying high latency is more important. Our team Introduced a new model with a continual learning concept that improved the efficacy of the false negative rate by more than 20% and accelerated the response time by three times. Besides, I Justified the continual learning technique's effectiveness and conducted multiple data visualizations to intuitively illustrate the research results.