演講公告

113/1/29 (一)13:30-15:00
成大資訊系 65104教室

Dr. Jinjun Xiong

Empire Innovation Professor with the Department of Computer Science & Engineering, University at Buffalo (UB)

MAKE EVERY BIT OF BANDWIDTH COUNT: AI SYSTEMS INNOVATION FOR ACCELERATING AI IN THE ERA OF CLOUD COMPUTING

Abstract

The computing landscape is undergoing significant shifts with AI's increasing integration into the cloud. Modern machine learning models demand extensive data access, posing challenges for programming and computing infrastructures. Managing data complexity across storage hierarchies while adapting to evolving infrastructures becomes crucial. This talk contends that many design options aren't optimal for AI workload acceleration on existing systems. Emphasizing the need for effective AI tools to identify system bottlenecks, it advocates for innovative hardware and software techniques to maximize system I/O bandwidth. This approach unveils fresh optimization avenues for accelerating AI workloads on the cloud. In addition, it advocates for innovative hardware and software techniques to maximize system I/O bandwidth, emphasizing the need for effective AI tools to identify system bottlenecks.

主辦單位：國立成功大學電資學院

協辦單位：智慧半導體及永續製造學院 Academy of Innovative Semiconductor and Sustainable Manufacturing　敏求智慧運算學院 Miin Wu School of Computing　IEEE Tainan Section