Keynotes

Joint Scheduling of Overlapping Phases in the MapReduce Framework

Professor Jie Wu
Director of the Center for Networked Computing and Laura H. Carnell Professor
Temple University, USA

Tuesday, 12 December 2017 | 09:30 - 10:30

Abstract: MapReduce includes three phases: map, shuffle, and reduce. Since the map phase is CPU-intensive and the shuffle phase is I/O-intensive, both can be conducted in parallel. This talk focuses on joint scheduling optimization of overlapping map and shuffle phases to minimize average job completion time. Challenges arise due to the dependency relationship between the map and shuffle phases since the shuffle phase relies on data emitted by the map phase. Our key observation is that the optimal schedule is trackable in two special cases. The first case is where the job set is pairable, i.e., the jobs are paired to avoid under-utilization of the I/O. The optimal schedule executes pairs of jobs depending on their size from smallest to largest. The second special case is where the job set is unilateral dominant, [KNW1] i.e., all jobs are simultaneously map-heavy, balanced, or shuffle-heavy. The optimal schedule executes the unilaterally dependent jobs in size order from smallest to largest. When the job set is a union of a pairable job set and a unilateral dominant [KNW2] job set, we propose an approximation schedule with a bound of 2 that resembles a mixture of the two optimal schedules above. Consequently, an arbitrary job set can be divided into two subsets to approximate the two special cases. Real data-driven experiments validate the efficiency and effectiveness of our algorithms.

Biography: Jie Wu is Director of the Center for Networked Computing and Laura H. Carnell professor at Temple University. He also serves as Director of International Affairs for the College of Science and Technology. He served as Chair of the Department of Computer and Information Sciences from the summer of 2009 to the summer of 2016 and as Associate Vice Provost for International Affairs from the fall of 2015 to the summer of 2017. Prior to joining Temple University, he was a program director at the National Science Foundation and was a distinguished professor at Florida Atlantic University. His current research interests include
mobile computing and wireless networks, routing protocols, cloud and green computing, network trust and security, and social network applications. Dr. Wu regularly publishes in scholarly journals, conference proceedings, and books. He serves on several editorial boards, including *IEEE Transactions on Service Computing* and the *Journal of Parallel and Distributed Computing*. Dr. Wu was general co-chair for IEEE MASS 2006, IEEE IPDPS 2008, IEEE ICDCS 2013, ACM MobiHoc 2014, ICPP 2016, and IEEE CNS 2016, as well as program co-chair for IEEE INFOCOM 2011 and CCF CNCC 2013. He was an IEEE Computer Society distinguished visitor, an ACM distinguished speaker, and the chair of the IEEE Technical Committee on Distributed Processing (TCDP). Dr. Wu is a CCF distinguished speaker and a fellow of the IEEE. He is the recipient of the 2011 China Computer Federation (CCF) Overseas Outstanding Achievement Award.
Industry Forecast: Cloudy with a 100% Chance of Science

Mr Mark Ryland
Director of Solutions Architecture and Chief Architect
World Wide Public Sector Team
Amazon Web Services, USA

Tuesday, 12 December 2017 | 11:00 - 12:00

Abstract: Hyper-scale cloud platforms continue to evolve at break-neck speed. Focusing on the Amazon Web Services (AWS) cloud, this talk first cover the latest technical innovations that make cloud computing more and more flexible and compelling to researchers (e.g., GPGPU and FPGA EC2 instance types; bare-metal EC2 and the Nitro architecture; advances in “serverless” computing, software-defined networking, storage services, and databases; etc.). We will briefly examine the explosion of new cloud services bringing artificial intelligence and machine learning (AI/ML) into the mainstream. We will discuss how the cloud’s security and access management features enable secure data-sharing and collaboration among researchers. Finally, we will review a number of interesting customer case studies that show why cloud computing is already at the leading edge of AI/ML as well as loosely-coupled HTC (high throughput computing), and is moving quickly up the performance and value chain for more traditional tightly-coupled HPC (high performance computing) use cases as well. With its extraordinary flexibility, speed of deployment, and pay-as-you-go pricing model, cloud computing is ushering in a new era of “experimental IT” that aligns with the mindset and the requirements of scientific researchers everywhere.

Biography: Mark Ryland is the technology leader for AWS Worldwide Public Sector (WWPS) team. Mr Ryland leads a team of technical professionals who provide architectural guidance, knowledge transfer, and implementation services to government and education customers around the globe. He also serves as a key interface between the WWPS team and the engineering, security, and compliance teams at AWS, ensuring that public sector customer requirements are front-and-center in cloud service planning and roadmaps. Ryland has more than 24 years of experience in the technology industry, including 13 years at Microsoft, the co-founding of two start-ups, and extensive work on technology standards and policy. His areas of technical expertise include networking, security, identity and access management, and distributed systems. He joined the AWS WWPS team as Chief Architect in September 2011, bringing a rich set of software engineering, distributed systems, cyber security, technical evangelism, and tech policy skills to the team.
Big Data Analytics and Intelligence at Alibaba

Dr Jingren Zhou
Vice President
Alibaba Group, China

Wednesday, 13 December 2017 | 09:00 - 10:00

Abstract: As the world’s largest e-commerce platform, Alibaba heavily relies on massive data analysis of many kinds to collect data insights and drive business decisions in real time. In this talk, I’ll present an overview of Big Data and AI computing platform at Alibaba, which consists of a wide range of products and services to enable fast and efficient big data development and intelligent analysis. The underlying computing infrastructure supports a variety of computation scenarios, including batch, interactive, stream, and graph computation, as well as large-scale machine learning on heterogeneous cloud-scale data centers. The platform not only supports Alibaba’s internal businesses but also provides solid services to enterprise customers via Alibaba Cloud. I’ll also describe key techniques and system internals, and outline outstanding research and engineering challenges.

Biography: Jingren Zhou is Vice President at Alibaba Group. He is responsible for driving Big Data development and business strategies at Alibaba Cloud Division. He manages the cloud engineering team to develop cloud-scale distributed computing platform, data analytic products, and various business solutions. He is also Head of the Search Division at Alibaba, leading the search engineering team to develop advanced techniques for personalized e-commerce and multimedia search, and provide best-in-class shopping experience at Alibaba’s e-commerce platforms, including Taobao and Tmall. Prior to joining Alibaba, Dr Zhou was partner engineering manager at Microsoft and researcher at Microsoft Research. His research interests include cloud-computing, distributed systems, and databases. Dr Zhou received his PhD in Computer Science from Columbia University.
Using Kernels to Harness the Complexity of Big Data Applications

Professor Benjamin W. Wah
Provost and Wei Lun Professor of Computer Science and Engineering
Chinese University of Hong Kong, Hong Kong

Thursday, 14 December 2017 | 09:00 - 10:00

Abstract: Big Data is emerging as one of the hottest multi-disciplinary research fields in recent years. Big data innovations are transforming science, engineering, medicine, healthcare, finance, business, and ultimately society itself. In this presentation, we examine the key properties of big data (volume, velocity, variety, veracity, and value) and their relation to some applications in science and engineering. To truly handle big data, new paradigm shifts will be necessary. Successful applications in big data will require in situ methods to automatically extracting new knowledge from big data, without requiring the data to be centrally collected and maintained. Traditional theory on algorithmic complexity may no longer hold, since the scale of the data may be too large to be stored or accessed. To address the potential of big data in scientific discovery, challenges on data complexity, computational complexity, and system complexity will need to be solved. We propose a new approach based on identifying kernels to harness the complexity of big data applications. Kernel data is a compact and manageable representation of the original data, with similar structure, data properties, or meta-properties. We illustrate these challenges and approaches by drawing on examples in various applications in science and engineering.

Biography: Benjamin W. Wah is currently the Provost and Wei Lun Professor of Computer Science and Engineering of the Chinese University of Hong Kong, as well as the Chair of the Research Grants Council of the University Grants Committee, Hong Kong, and the Franklin W. Woeltge Emeritus Professor of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign. Before then, he served as the Director of the Advanced Digital Sciences Center in Singapore, as well as the Franklin W. Woeltge Professor of Electrical and Computer Engineering and Professor of the Coordinated Science Laboratory of the University of Illinois, Urbana-Champaign, IL. He received his Ph.D. degree in computer science from the University of California, Berkeley, CA, in 1979. He has received numerous awards for his contributions, which include the IEEE CS Technical Achievement Award (1998), the IEEE Millennium Medal (2000), the IEEE-CS W. Wallace-McDowell Award (2006), the Pan Wen-Yuan Outstanding Research Award (2006), the IEEE-CS Richard E. Merwin Award (2007), the IEEE-CS Tsutomu Kanai Award (2009), and the Distinguished Alumni Award in Computer Science of the University
of California, Berkeley (2011). Wah's current research interests are in the areas of big data applications and multimedia design and processing.

Wah cofounded the *IEEE Transactions on Knowledge and Data Engineering* in 1988 and served as its Editor-in-Chief between 1993 and 1996. He currently serves as the Honorary Editor-in-Chief of *Knowledge and Information Systems* and is on the editorial boards of *Information Sciences, International Journal on Artificial Intelligence Tools, Journal of VLSI Signal Processing, World Wide Web,* and *Journal of Computer Science and Technology.* He has served the IEEE Computer Society in various capacities, including Vice President for Publications (1998 and 1999) and President (2001). He is a Fellow of the AAAS, ACM, and IEEE.