ParLearning 2015 Keynote Speaker 1

Professor David Bader, Georgia Institute of Technology, USA

Title: Massive-Scale Analytics

Abstract:
Emerging real-world graph problems include: detecting community structure in large social networks; improving the resilience of the electric power grid; and detecting and preventing disease in human populations. Unlike traditional applications in computational science and engineering, solving these problems at scale often raises new challenges because of the sparsity and lack of locality in the data, the need for additional research on scalable algorithms and development of frameworks for solving these problems on high performance computers, and the need for improved models that also capture the noise and bias inherent in the torrential data streams. In this talk, the speaker will discuss the opportunities and challenges in massive data-intensive computing for applications in computational science and engineering.

***Bio*:**
Dr. David Bader is a Full Professor and Chair of the School of Computational Science and Engineering, College of Computing, at Georgia Institute of Technology, and Executive Director of High Performance Computing. He received his Ph.D. in 1996 from The University of Maryland, and his research is supported through highly-competitive research awards, primarily from NSF, NIH, DARPA, and DOE. Dr. Bader serves as a board member of the Computing Research Association (CRA), on the NSF Advisory Committee on Cyberinfrastructure, on the Council on Competitiveness High Performance Computing Advisory Committee, on the IEEE Computer Society Board of Governors, and on the Steering Committees of the IPDPS and HiPC conferences. He is the editor-in-chief of IEEE Transactions on Parallel and Distributed Systems (TPDS) and Program Chair for IPDPS 2014. Bader also serves as an associate editor for several high impact publications including IEEE Transactions on Computers (TC), ACM Transactions on Parallel Computing (TOPC), and ACM Journal of Experimental Algorithmics (JEA).

978-1-4673-7684-6/15 $31.00 © 2015 IEEE
DOI 10.1109/IPDPSW.2015.179

ParLearning 2015 Keynote Speaker 2

Professor Yihua Huang, Nanjing Univesity, China

Title: Unified Programming Model and Platform for Big Data Machine Learning and Data Analytics

Abstract:

Machine learning is among the core techniques for big data analytics. Nowadays, there are diversified parallel programming models and computing platforms such as Hadoop MapReduce, Spark, MPI, CUDA-GPU, etc. In terms of data sizes and specific problems, we can choose one of these to implement our parallel machine learning algorithms for analyzing big data. However, when such a new model and platform comes up, we need to re-write all algorithms we use, which is a lot of duplicate work and burden. On the other hand, for data analysts from different application domains who focus on high-level business modeling and data analytics by using high-level data analytic programming language and tool such as R or Matlab, it is really hard for them to learn and use the relatively low-level parallel programming models and computing platforms.

This talk will present a number of commonly-used parallel machine learning algorithms implemented with Hadoop MapReduce and Spark. Then a unified programming model and platform for machine learning and data analytics across heterogeneous parallel programming models and computing platforms will be presented. The key idea is that we will adopt the large-scale matrix as a unified abstraction for representing different machine learning and data analytic algorithms. Then we adopt R as the high-level programming language. We design and provide a unified framework and platform, Octopus, that take charge of converting R and mapping large-scale matrix computation into underlying parallel programming models and platforms. Based on the unified model and platform, we can achieve the goal of "write once, run anywhere" for big data machine learning and data analytic algorithms.

***Bio*:**

Dr. Yihua Huang is currently a professor in the Department of Computer Science and Technology at Nanjing University in China, Vice Secretary General of the Big Data Committee of the China Computer Society(CCF), and Chair of the Big Data Committee of the Jiangsu Province Computer Society. He received his bachelor, master and Ph.D. degree in computer science from Nanjing University in 1983, 1986 and 2007 respectively. His main research interests include big data processing, parallel and distributed computing, semantic analytics, web data mining. He is one of the earliest researchers in China on big data processing. Currently he is engaged in a series of researches on big data processing with research grants from the state research programs in China as well as grants from industry partners including Intel, Google, ZTE and Baidu. In 2009, sponsored by Google, he created and offered the course "MapReduce Big Data Parallel Processing" for graduate students in computer science at Nanjing University and received the "2012 Google Faculty Fellowship" from Google in 2012.

ParLearning 2015 Keynote Speaker 3

Professor Ananth Kalyanaraman, Washington State University, USA

Title: Problems, Challenges and Opportunities in Exploring the "Dark Matter" of Life Sciences: The Microbiome

Abstract:
Often labeled as the "dark matter" of the life sciences, the term microbiome refers to the universe of microbes that inhabit the various environment habitats that surround us and present within us. Yet we know very little about this microbial universe. Rapid advances in high-throughput sequencing and other technologies over the decade are changing that. New data from previously unsequenced habitats are being sampled to previously unprecedented depths. This has aptly led to an immense interest among computer scientists and biologists to combine their expertise in pursuit of developing new tools and implementing new functions geared towards acquiring fundamental insights into the mechanics of microbial machinery. In this talk, we will visit problems, challenges and an outlook on the promises and opportunities of this emerging branch within computational life sciences. More specifically, we will focus on the computational techniques that have been developed and their applications. What we are yet to learn from these microbial communities could have a significant impact on how we shape our future.

***Bio*:**
Ananth Kalyanaraman's main research interest is in building high performance computational tools for solving data-driven and data-intensive problems within the life sciences. He is an Associate Professor and Boeing Centennial Chair in Computer Science at the School of Electrical Engineering and Computer Science in Washington State University. He received his Bachelor of Engineering from Visvesvaraya National Institute of Technology in Nagpur (India) in 1998, and his PhD from Iowa State University in 2006. Ananth is a recipient of a DOE Early Career Award, Early Career Impact Award from Iowa State University, and two best paper awards. He has organized several workshops and mini-symposia relating to high performance computational biology and Big Data techniques for the life sciences at IEEE, ACM and SIAM conferences. Ananth is a member of AAAS, ACM, IEEE-CS, and ISCB.