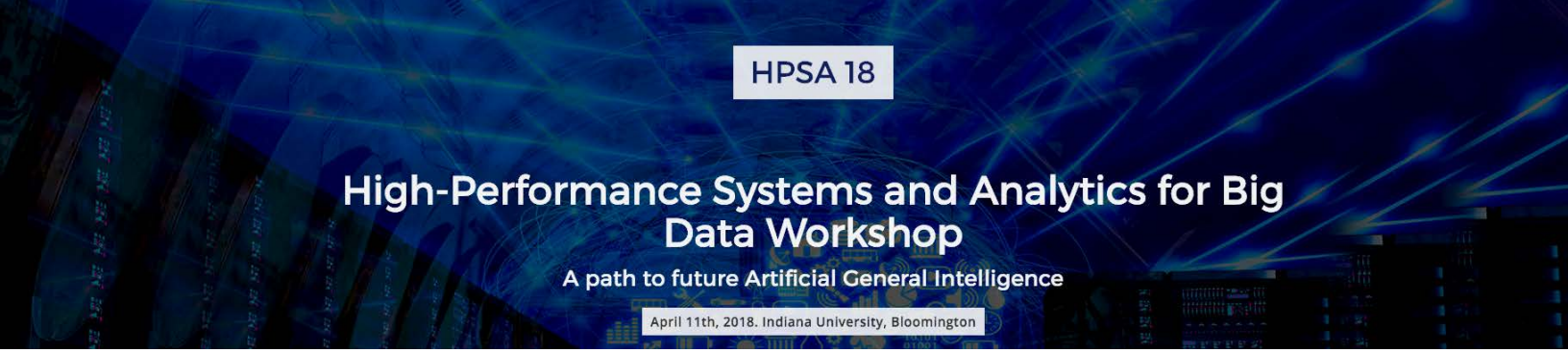


Message from Workshop Chairs

The workshop chairs welcome the distinguished international participants to this short workshop on the intersection of Data Analytics and High-Performance Computing. We will produce a report based on the talks and Panel discussion that will hopefully guide new research in this area.

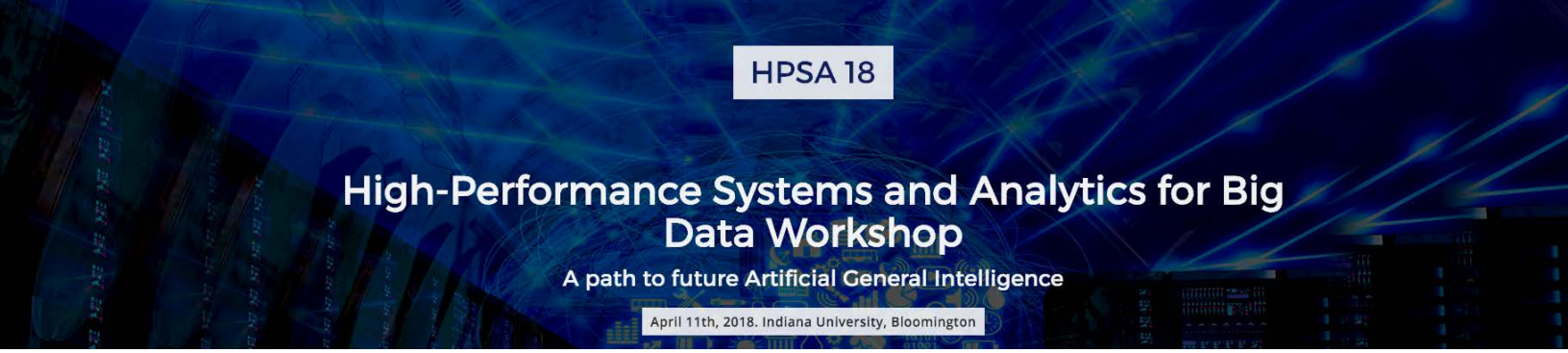
Prof. X. Sean Wang, Dean of the School of Computer Science, Fudan University
Prof. Judy Qiu, Intelligent Systems Engineer, SICE, Indiana University

Organizing Committee:
Julie Overfield, Assistant Director, ISE
Gina Marie Gallagher, Senior Director of Corporate & Foundation Relations, SICE
Whitney Riley, Executive Director, Corporate Relations, OVPR

Workshop Abstract

The workshop theme is “High-Performance Systems and Analytics for Big Data” (HPSA). It involves Large-Scale Data Analytics on High-Performance Computing (HPC) clusters optimized for data analysis. HPSA has been identified by Gartner in their infrastructure strategies priority matrix under the rubric of Hyperscale computing, as having transformational importance with their top rating in the 5-10 year timeframe. The potential impact on scientific discovery and economic development from Data Analytics is tremendous.

The workshop features innovative research and development in hardware, algorithms and software for big data systems of transformational capability on computer architectures ranging from commodity clouds, hybrid HPC-clouds, and supercomputers. It aims at performance and security that scales and fully exploit the specialized features (communication, memory, energy, I/O, accelerator) of each different architecture. Studies of new architectures and benchmarking of existing systems will be covered. Applications will range from pleasingly parallel, MapReduce, to Machine Learning (e.g., Random Forest, SVM, Latent Dirichlet Allocation, Clustering and Dimension Reduction), Deep Learning, and Large Graph Analytics.



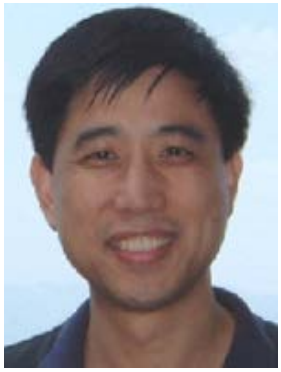
Workshop Agenda

7:00 am – 8:00 am	**Breakfast**	Luddy Room 3166 (3rd floor)
8:00 am – 8:10 am	**Opening Remarks** Prof. Rick Van Kooten, Vice Provost for Research	IU
8:10 am - 8:20 am	Judy Qiu	IU
8:20 am – 8:30 am	X. Sean Wang	Fudan
8:30 am – 9:00 am	Linton Ward	IBM (Keynote$_1$)
9:00 am – 9:20 am	Nathan Greeneltch	Intel
9:20 am – 9:40 am	Takuya Araki	NEC
10:00 am – 10:30 am	**Break**	
10:30 am – 10:50 am	Anthony Skjellum	UTC
10:50 am – 11:10 am	Andrew Younge	Sandia
11:10 am – 11:30 am	Anil Vullikanti	Virginia Tech
11:30 am – 11:50 am	Albert Jonathan	Univ. of Minnesota
12:00 pm – 1:00 pm	**Lunch**	Community Center (1st floor)
1:10 pm – 1:40 pm	Tony Hey	UK STFC (Keynote$_2$)
1:40 pm – 2:00 pm	Piotr Luszczek	Univ. of Tenessee
2:00 pm – 2:20 pm	Scott Michael	UITS
2:20 pm – 2:40 pm	Wo Chang	NIST
2:40 pm – 3:00 pm	**Break**	
3:00 pm – 3:20 pm	X. Sean Wang	Fudan
3:20 pm – 3:40 pm	Weihua Zhang	Fudan
3:40 pm – 4:00 pm	Martin Swany	IU
4:00 pm – 4:20 pm	Geoffrey Fox	IU
4:30 pm – 6:00 pm	**Panel Session (Chair: Dennis Gannon)** X. Sean Wang, Tony Hey, Wo Chang, Linton Ward, Nathan Greeneltch, Takuya Araki, Judy Qiu	Community Center (1st floor)
6:30 pm – 8:30 pm	**Dinner**	Finch's Brasserie Restaurant

Speaker Information

Prof. X. Sean Wang, Dean of the School of Computer Science, Fudan University

Bio



Prof. X. Sean Wang

is Professor and Dean at the School of Computer Science, Fudan University, Shanghai, China. He received his PhD degree in Computer Science from the University of Southern California, Los Angeles, California, USA. Before joining Fudan University in 2011, he was the Dorothean Chair Professor in Computer Science at the University of Vermont, Burlington, Vermont, USA, and between 2009-2011, he served as a Program Director at the National Science Foundation, USA, in the Division of Information and Intelligent Systems. He has published widely in the general area of databases and information security and was a recipient of the US National Science Foundation Research Initiation and CAREER awards. His research interests include database systems, information security, and data mining. In the research community, he served as the general chair of IEEE ICDE 2011 held in Washington DC and ACM CIKM 2014 in Shanghai, China, and in various other roles at international conferences and journals. He's currently chief editor of the Springer Journal of Data Science and Engineering, associate

editor of IEEE Transactions on Cloud Computing, and of Geoinformatica and WWW journal, and past associate editor of TKDE and KAIS. He's also currently on the steering committees of the IEEE ICDE and IEEE BigComp conference series, and past Chair of WAIM Steering Committee.

Title: Building the cognitive platform to accelerate innovation (Keynote 1)

By Dr. Linton Ward, Distinguished Engineer, IBM

Abstract

We are in a period of dramatic innovation in the use of analytics in nearly every discipline leading to many opportunities as well as disruptions. Leading innovators are adopting the cognitive platform to support their artificial intelligence activities. This innovation is enabled by a number of underlying trends including vast amount of available digital data and open innovation in both software and hardware: Python and PyData communities, accelerated and hybrid computing models. This talk will describe the challenges teams face in creation of the cognitive platform and then describe the combinations of elements to address those challenges.

The open data science workbench is the means to enable an AI workflow and data flow, and the Artificial Intelligence Grid is a means to support the computation needs of a team of data scientists and developers. Supporting this grid is essentially an HPC cluster that supports accelerated computation, high performance data movement, and fast working storage and big data persistent storage. These technologies enable new science, but also present research opportunities as well. This talk will close with a short discussion on the opportunities to leverage breakthroughs in latency, bandwidth and coherence among accelerators and CPU memories.

Bio

Dr. Linton Ward

is an IBM Distinguished Engineer leading open and cognitive workload solution development on OpenPower systems. He is passionate about designing and building relevant and game-changing solutions to tackle business and HPC computing challenges. He has designed and optimized many hardware-software stack solutions and led hardware design teams for numerous integrated offerings, including Data Warehouse, Hadoop, analytics and HPC offerings. As a systems and solutions architect, Linton brings a unique combination of deep understanding of software needs, client experience and hardware capability. He regularly meets with clients to help them understand the solution space and help them define the next steps in their analytic journey.

Title: Learn Faster with Intel Data Analytics Acceleration Library (DAAL)

By Dr. Nathan Greeneltch, Intel Corporation

Abstract

This talk with give an overview of the Intel DAAL library, complete with available features and performance charts. It will announce availability of the new features, including those in future beta releases. A deep dive into a select learning algorithm is planned, highlighting the programming advances inside this library. Intel DAAL includes optimization of each portion of the analytical workflow and includes C++, Python, and Java APIs.

Bio

Dr. Nathan Greeneltch

joined the Technical Computing, Analyzers and Runtimes (TCAR) group in 2017 as a technical consulting engineer (TCE). His role is to help drive customer engagements for Python as well as Intel's libraries, leveraging the synergies between Python and MKL. Before joining the TCAR team, Nathan spent 3 years in the processor development side of Intel where he was a ML practitioner in the defects division, identifying and predicting failure areas in the coming generations of Intel processor. Nathan has a PhD in physical chemistry from Northwestern University, where he worked on nanoscale lithography of metal wave-guides for amplification of laser-initiated vibrational signal in small molecules.

Title: SX-Aurora TSUBASA and its application to machine learning

By Dr. Takuya Araki, NEC Corporation

Abstract

NEC has released a new vector computer SX-Aurora TSUBASA. Traditional vector computers have been utilized mainly for HPC. However, we are extending application area of SX-Aurora TSUBASA with its wide range of product lineup that starts from small scale machines. In this talk, we will introduce its application to AI / machine learning. Vector computers are faster than general purpose CPUs because they load, calculate, and store many elements of data (e.g. 256) at a time in parallel. In addition, X-Aurora TSUBASA provides very high memory bandwidth (1.22 TB/s) to attain high efficiency by feeding enough data to ALU and FPU. On the other hand, large scale machine learning tends to utilize sparse matrix, which requires high memory bandwidth. Therefore, such work load is promising for SX-Aurora TSUBASA. We implemented middleware on SX-Aurora TSUBASA and realized distributed and vectorized machine learning on top of it. The evaluation results show that up to 100 times speed up could be attained compared to general purpose CPU with existing middleware. In addition, we extended Spark, which is commonly used middleware for statistical machine learning, to utilize the vector processor through our middleware; machine learning algorithms are offloaded to SX-Aurora TSUBASA and users can enjoy its high performance without knowing the hardware details.

Bio

Dr. Takuya Araki

received the B.E., M.E., and Ph.D. degrees from the University of Tokyo, Japan in 1994, 1996, and 1999, respectively. He was a visiting researcher at Argonne National Laboratory from 2003 to 2004. He is currently a principal researcher of system platform research laboratories, NEC Corporation. His research interests include parallel and distributed computing, big data analytics, and multimedia information retrieval. He is a director of Information Processing Society of Japan (2017-2018).

Title: Finding Trees and Anomalous Subgraphs in Parallel

By Prof. Anil Vullikanti, Computer Science and the Biocomplexity Institute, Virginia Tech

Abstract

We focus on two classes of problems in graph mining here: (1) finding trees and (2) anomaly detection using network scan statistics in complex networks, which involves finding connected subgraphs that maximize a

suitable anomaly score depending on the underlying application. These are fundamental problems in a broad class of applications. Most of the parallel algorithms for such problems are either based on heuristics, which do not scale very well, or use techniques like color coding, which have a high memory overhead. We describe a novel approach for parallelizing both these classes of problems,

using algebraic representations for such problems---this involves detecting multilinear terms in multivariate polynomials. Our algorithms show good scaling over a large regime, and run on networks with close to half a billion edges. The resulting parallel algorithm for trees is able to scale to subgraphs of size 18, which has not been shown before, and significantly outperforms the best prior color coding based methods. Our algorithm for network scan statistics is the first such parallelization, and is able to handle a broad class of scan statistics (both parametric and non-parametric), with the same approach.

Bio

Prof. Anil Vullikanti

is an Associate Professor in the Department of Computer Science and the Biocomplexity Institute of Virginia Tech. His interests are in the areas of approximation and randomized algorithms, distributed computing, graph

dynamical systems and their applications to epidemiology, social networks and wireless networks. He is a recipient of the NSF and DOE Career awards.

Title: Geo-Distributed Clouds For Data Analytics

By Prof. Jon Weissman, University of Minnesota

Abstract

Data is increasingly being created at the network edge in many domains of interest. Further, in many domains (e.g. CDN log analysis) data is generated in a geo-distributed fashion at many locations including edge nodes and data centers alike. In this realm, high performance in-situ data analytics is needed to process both rapidly arriving streaming data as well as stored batch-oriented data. We describe the motivation and challenges in performing both kinds of geo-distributed edge analytics and our solution, called Nebula that has been applied to a wide-variety of data generation contexts.

Bio

Prof. Jon Weissman (presented by Albert Jonathan)

is a Professor of Computer Science at the University of Minnesota where he co-leads the Distributed Computing Systems Group. His research interests

are in distributed systems, cloud/edge computing, high performance computing, and storage systems.

Title: Big Scientific Data and Data Science (Keynote 2)

By Prof. Tony Hey, Rutherford Appleton Laboratory, UK

Abstract

There is broad recognition within the scientific community that the ongoing deluge of scientific data is fundamentally transforming academic research. "The Fourth Paradigm" refers to the new 'data intensive science' and the tools and technologies needed to manipulate, analyze, visualize, and manage large amounts of research data. This talk will review the challenges posed by the growth of 'Experimental and Observational Data' (EOD) generated by the new generation of large-scale experimental facilities at the UK's Harwell site near Oxford. The talk will conclude with a discussion of the use of experimental scientific 'benchmarks' both for training the scientist users of these facilities in Machine Learning and data science technologies and as a vehicle for research in the robustness and transparency of the predictions of these algorithms.

Bio

Prof. Tony Hey

began his career as a theoretical physicist with a doctorate in particle physics from the University of Oxford in the UK. After a career in physics that included research positions at Caltech and CERN, and a professorship at the University of Southampton in England, he became interested in parallel computing and moved into computer science. In the 1980's he was one of the pioneers of distributed memory message-passing computing and co-wrote the first draft of the successful MPI message-passing standard.

After being both Head of Department and Dean of Engineering at Southampton, Tony Hey escaped to lead the U.K.'s ground-breaking 'eScience' initiative in 2001. He recognized the importance of Big Data for science and wrote one of the first papers on the 'Data Deluge' in 2003. He joined Microsoft in 2005 as a Vice President and was responsible for Microsoft's global university research engagements. He worked with Jim Gray and his multidisciplinary eScience research group and edited a tribute to Jim called 'The Fourth Paradigm: Data-Intensive Scientific Discovery.' Hey left Microsoft in 2014 and spent a year as a Senior Data Science Fellow at the eScience Institute at the University of Washington. He returned to the UK in November 2015 and is now Chief Data Scientist at the Science and Technology Facilities Council.

Title: HPC Autotuning Techniques for Computational Kernels in Data Analytics

By Prof. Piotr Luszczek, University of Tennessee

Abstract

The increasing demand for computational power to analyze scientific data creates new challenges for traditional HPC techniques of arriving at optimized code variants. In this talk I will present our approach that automates various aspects of deriving code sequence that guarantee high

performance in a wide range of aspects of data analytics workloads. The example applications where we successfully use our methods include machine learning techniques for image registrations and image classifiers based on deep neural networks.

Bio

Prof. Piotr Luszczek

is a Research Assistant Professor at the University of Tennessee and a research director in the Innovative Computing Laboratory in Knoxville. He holds a PhD in Computer Science (Advanced Performance Optimizations for Sparse Direct and Iterative Methods) from the University of Tennessee. His core research is centered on performance modeling and evaluation as well as parallel programming paradigms and languages.

Title: Supporting High Performance Analysts with System Software for Virtualized Supercomputing

By Dr. Andrew J. Younge, Sandia National Laboratories

Abstract

In just the past few years, data science has provided critical methods for gaining scientific knowledge and enabling novel discoveries across a wide range of disciplines. As this data deluge grows ever greater, large-scale data analytics and machine learning workloads are quickly becoming critical computational tools within the scientific community. This includes analytics operations not only as a processing component to large High-Performance Computing (HPC) simulations, but also as standalone scientific tools for knowledge discovery to handle a world that is increasingly more instrumented and interconnected. Scientific computing is at the verge of convergence between HPC and big data analytics workloads, however, current big data software stacks often struggle to leverage the investments made in supercomputing within the U.S. DOE.

This talk will describe some of the unique efforts at Sandia National Laboratories for tackling the convergence between big data, machine learning, and HPC. We will introduce some the research activities underway within scalable system software at Sandia, as well as identify potential avenues for future collaboration. The goal of this effort is to enable the support of diverse ecosystems within a scalable architecture that enable all distributed computational workloads, beyond just traditional bulk synchronous parallel jobs.

Bio

Dr. Andrew J. Younge

is a R&D Computer Scientist at Sandia National Laboratories with the Scalable System Software group. His research interests include the intersection of High Performance Computing, Cloud Computing, Distributed Systems, and energy efficient computing. Andrew has a Ph.D in Computer Science from Indiana University, where he was a Persistent Systems fellow and a member of the FutureGrid project. Over the years, Andrew has held visiting positions at the MITRE Corporation, the University of Southern California / Information Sciences Institute, and the University of Maryland, College Park. He received his B.S. and M.S from the Computer Science Department at Rochester Institute of Technology (RIT) in 2008 and 2010, respectively. During this time, Andrew worked as a Graduate Researcher on the Cyberaide project and as a research assistant on an experimental Social Psychology research project.

Title: NIST PWG Big Data Reference Architecture for HPC and Analytics

By Mr. Wo Chang, NIST

Abstract

Big Data is the term used to describe the deluge of data in our networked, digitized, sensor-laden, information driven world. There is a broad agreement among commercial, academic, and government leaders about the remarkable potential of “Big Data” to spark innovation, fuel commerce, and drive progress. The availability of vast data resources carries the potential to answer questions previously out of reach. However, there is also broad agreement on the ability of Big Data to overwhelm traditional approaches.

Big Data architectures come in many shapes and forms ranging from academic research settings to product-oriented workflows. With massive-

scale dynamic data being generate from social media, Internet of Things, Smart Cities, and others, it is critical to analyze these data in real-time and provide proactive decision. With the advancement of computer architecture in multi-cores and GPUs, and fast communication between CPUs and GPUs, parallel processing utilizes these platforms could optimize resources at a reduced time. This presentation will provide the past, current, and future activities of the NIST Big Data Public Working Group (NBD-PWG) and how the NIST Reference Architecture may address the rate at which data volumes, speeds, and complexity are growing requires new forms of computing infrastructure to enable Big Data analytics interoperability such that analytics tools can be re-usable, deployable, and operational.

The focus of NBD-PWG is to form a community of interest from industry, academia, and government, with the goal of developing consensus definitions, taxonomies, secure reference architectures, and standards roadmap which would create vendor-neutral, technology and infrastructure agnostic framework. The aim is to enable Big Data stakeholders to pick-and-choose best analytics tools for their processing under the most suitable computing platforms and clusters while allowing value-additions from Big Data service providers and flow of data between the stakeholders in a cohesive and secure manner.

Bio

Mr. Wo Chang

is Digital Data Advisor for the NIST Information Technology Laboratory (ITL). His responsibilities include, but are not limited to, promoting a vital and growing Big Data community at NIST with external stakeholders in commercial, academic, and government sectors. Mr. Chang currently chairs the ISO/IEC JTC 1/WG 9 Working Group on Big Data, IEEE Big Data Governance and Metadata Management, and NIST Big Data Public Working Group. Prior to joining ITL Office, Mr. Chang was manager of the Digital Media Group in ITL and his duties included overseeing several key projects including digital data, long-term preservation and management of EHRs, motion image quality, and multimedia standards. In the past, Mr. Chang was the Deputy Chair for the US National Body for MPEG (INCITS L3.1) and chaired several other key projects for MPEG, including MPQF, MAF, MPEG-7 Profiles and Levels, and co-chaired the JPEG Search project. Mr. Chang was one of the original members of the W3C's SMIL WG and developed one of the SMIL reference software. Furthermore, Mr. Chang also participated in the HL7 and ISO/IEC TC215 for health informatics and IETF for the protocols development of SIP, RTP/RTPC, RTSP, and RSVP networking protocols. Mr. Chang's research interests include, big data analytics, high performance and cloud computing, content metadata description, digital file formats, multimedia synchronization, digital data preservation, and Internet protocols.

Title: The Data placement and transformation in Astronomy data processing

By Prof Wei Wang, Fudan University

Abstract

In this talk we discuss the challenge in the astronomy data process applications such as SKA data process system first. One challenge in

SKA/SDP is that there is a huge data transformation cost in the system. Most of the transformation cost comes from the Join operation in Spark. We present the cost model for the astronomy data processing, and then present several data process rewriting approaches for Join operation on SPARK to improve the performance of join operations for SDP.

Bio

Prof. Wei Wang (presented by Prof. Sean Wang)

is a Professor in the School of Computing Science at Fudan University. He received his PhD in Computer Science in 1998 from the Department of Computer Science in the Fudan. He won the second prize of Natural Science Award by the Ministry of Education of China in 2012. He was grantee of the New Century Excellent Talents program of the Ministry of Education of China and the Shanghai Science and Technology Committee Rising-Star Program in 2005. He has served as a PC member and reviewer for major conferences (e.g., ICDM, CIKM, WAIM). His research area includes data management and analysis in IOT, knowledge graph, medical information management. He has published more than 30 papers in the major conferences and journals (e.g., SIGMOD, VLDB, SIGKDD, CIKM, ICDE, ICDM, TKDE).

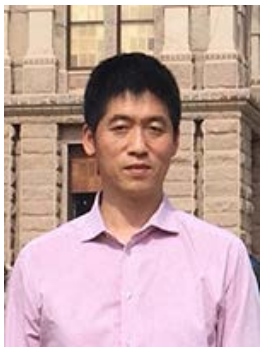
Title: Eunomia: Scaling Concurrent Search Trees under Contention Using HTM

By Prof Weihua Zhang, Fudan University

Abstract

While hardware transactional memory (HTM) has recently been adopted to construct efficient concurrent search tree structures, such designs fail to deliver scalable performance under contention. In this talk, I first conduct a detailed analysis on an HTM-based concurrent B+Tree, which uncovers several reasons for excessive HTM aborts induced by both false and true conflicts under contention. Based on the analysis, I will introduce Eunomia, a design pattern for search trees which contains several principles to reduce HTM aborts, including splitting HTM regions with version- based concurrency control to reduce HTM working sets, partitioned data layout to reduce false conflicts, proactively detecting and avoiding true conflicts, and adaptive concurrency control. To validate their effectiveness, such designs are applied to construct a scalable concurrent B+Tree using HTM. Evaluation using key-value store benchmarks on a 20-core HTM-capable multi-core machine shows that Eunomia leads to 5X-11X speedup under high contention, while incurring small overhead under low contention.

Bio



Prof. Weihua Zhang

is a Professor in the School of Computing Science at Fudan University. He received his PhD in Computer Science in 1998 from the Department of Computer Science in the Fudan. He won the second prize of Natural Science Award by the Ministry of Education of China in 2012. He was grantee of the New Century Excellent Talents program of the Ministry of Education of China and the Shanghai Science and Technology Committee Rising-Star Program in 2005. He has served as a PC member and reviewer for major conferences (e.g., ICDM, CIKM, WAIM). His research area includes data management and analysis in IOT, knowledge graph, medical information management. He has published more than 30 papers in the major conferences and journals (e.g., SIGMOD, VLDB, SIGKDD, CIKM, ICDE, ICDM, TKDE).

Title: Hardware-Accelerated Network Microservices for Big Data and Extreme Scale Computing

Prof. Martin Swany, Indiana University

Bio

Prof. Martin Swany

is Associate Chair and Professor in the Intelligent Systems Engineering Department in the School of Informatics, Computing, and Engineering at Indiana University. His research interests include embedded systems and reconfigurable computing as well as high-performance parallel and distributed computing and networking.

Title: High Performance Big Data Computing in the Digital Science Center

By Prof. Geoffrey Fox, Department Chair and Interim Associate Dean for Intelligent Systems Engineering, Indiana University

Abstract

As a collaboration funded by the NSF Dibbs program Qiu and Fox are developing the components that can be used to build Big Data Analysis Systems with scalable HPC performance and the functionality of ABDS – the Apache Big Data Software Stack. One highlight is a novel HPC-Cloud convergence framework Harp-DAAL with a kernel Machine Learning library exploiting the Intel node library DAAL and HPC communication collectives within the Hadoop ecosystem. The broad applicability of Harp-DAAL is supporting all 5 classes of data-intensive computation, from pleasingly parallel to machine learning and simulations. Another highlight is Twister2 which consists of a set of middleware components to support batch or streaming data capabilities familiar from Apache Hadoop, Spark, Heron and Flink but with high performance. Twister2 covers bulk synchronous and data flow communication; task management as in Mesos, Yarn and Kubernetes; dataflow graph execution models; launching of the Harp-DAAL library; streaming and repository data access interfaces, in-memory databases and fault tolerance at dataflow nodes. These are available in current Apache systems but as integrated packages which do not allow needed customization for different application scenarios.

Bio

Prof. Geoffrey Fox

is a distinguished professor of Engineering, Computing, and Physics at Indiana University where he is director of the Digital Science Center, and Department Chair and Interim Associate Dean for Intelligent Systems Engineering at the School of Informatics, Computing, and Engineering. He previously held positions at Caltech, Syracuse University, and Florida State University after being a postdoc at the Institute for Advanced Study at Princeton, Lawrence Berkeley Laboratory, and Peterhouse College Cambridge. He has supervised the Ph.D. of 70 students and is a Fellow of APS (Physics) and ACM (Computing) and works on the interdisciplinary interface between computing and applications.

Prof. Dennis Gannon, Indiana University

Bio

Prof. Dennis Gannon

is Professor Emeritus in the School of Informatics and Computing at Indiana University. From 2008 until he retired in 2014 Dennis Gannon was with Microsoft Research, most recently as the Director of Cloud Research Strategy. His previous roles at Microsoft include directing research as a member of the Cloud Computing Research Group and the Extreme Computing Group. From 1985 to 2008 Gannon was with the Department of Computer Science at Indiana University where he was Science Director for the Indiana Pervasive Technology Labs and, for seven years, Chair of the Department of Computer Science. His research interests include large-scale cyberinfrastructure, programming systems and tools, distributed computing, machine learning and data analytics and parallel programming He has published over 200 scientific articles and 4 books. He has a Ph.D. in Computer Science from the University of Illinois, Urbana Champaign and a Ph.D in Mathematics from the University of California, Davis.

Panel session (Chair: Prof. Dennis Gannon)

1. What are the important difficult big data problems that are yet be solved? Where are solutions to these problems needed? Is there a specific grand challenge that we can solve in the next 5 to 10 years? If so, what is it?

2. For high-performance big data analytics, the fastest methods are on HPC clusters. But can we make them work in dynamic interactive data analysis? Also, Edge computing is now an essential tool for a world of streaming data and device control. The cloud companies are devoting vast resources to making the edge work seamlessly with the cloud data centers. Can we make the new HPC-based methods work across edge-cloud-and-HPC seamlessly?

3. Cloud data centers are evolving supercomputing like features to support big data and machine learning. This includes Google's Tensorflow chip and Azure's FPGA configurable mesh and hardware microservices. Supercomputer centers are now embracing container technologies like Singularity and they are toying with microservice fabrics like Kubernetes. Will we ever see a true convergence of supercomputing and cloud?

4. What are the data science research and application challenges in both industry and science that are not being addressed by current trends? Is security a key problem?

5. If we have examples from the previous question, how will we fund work on them? It seems likely that China's government will make the needed investments, but what about in the US? Europe?

6. What is the measure of success in 5-10 years?

How Do You Get to Indiana University?

Option 1

Shuttle bus at Zone 1 of the Ground Transportation Center. Indiana Memorial Union (IMU) is a bus stop in Bloomington.

Option 2

Online Reservation of Classic Touch Limousine Service or Toll Free: 800.319.0082

Option 3

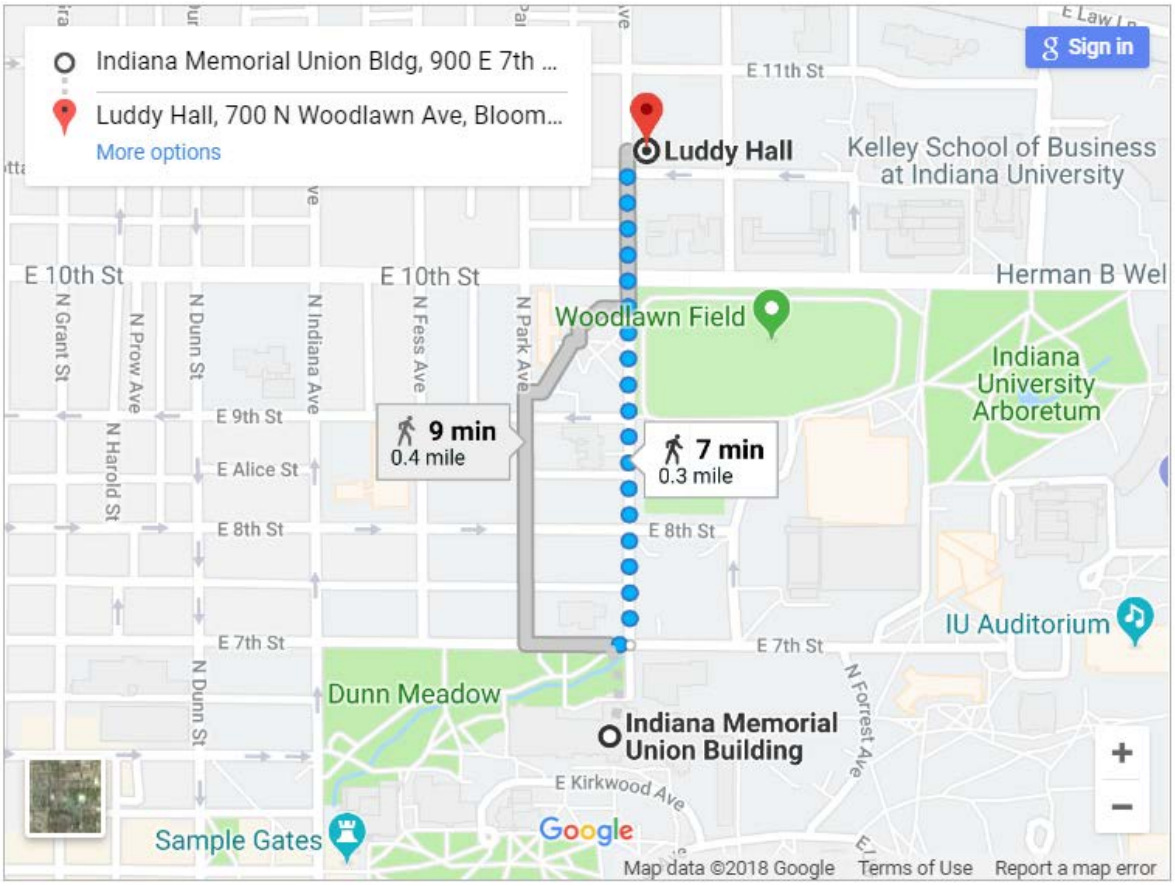
Ride with Uber at Zone A of the Ground Transportation Center.

Option 4

Car rental at the Indianapolis Airport.

How Do You Get to Luddy Hall from IMU?

The HPSA workshop will be hosted in the Executive Conference Room 3166 of Luddy Hall. It's a short walking distance from the Briddle Hotel in IMU to the Luddy Hall (see map below).



Workshop Report