2016
IEEE International Conference
on
Big Data

December 5 - December 8, 2016 • Washington DC, USA

Sponsored by

National Science Foundation
WHERE DISCOVERIES BEGIN

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

2016 IEEE International Conference on Big Data

IEEE Big Data 2016 Program Schedule

Washington DC, USA
December 5 - December 8, 2016

Keynote Lecture: ***60 minutes*** (about 45 minutes for talk and 15 minutes for Q and A)
Main conference regular paper: ***25 minutes*** (about 20 minutes for talk and 5 minutes for Q and A)
Main conference short paper: ***15 minutes*** (about 11 minutes for talk and 4 minutes for Q and A)

All conference activities take place at the Hyatt Regency Washington on Capitol Hill.

Sunday, 4-December			
3:00 – 8:00 pm Location:	**Registration** *Regency Foyer*		

Monday, 5-December			
7:20am-6:00 pm Location:	**Registration** *Regency Foyer*		
10:00-10:20 am and 3:30 – 3:50 pm Location:	**Coffee Break** *Regency Foyer*		
2:00 – 6:00 pm Location:	**Poster Session (Set up only)** *Regency Foyer and Hall of Battles*		
Time	**Session/Workshops**	**Session Chair**	**Location**
Full Day Workshops 8:00 – 6:30 pm	**W2**, 2nd International workshop on Big Data for sustainable Development	Dr. Aki-Hiro Sato	*Regency A*
	W3&W14, 3rd Workshop on Advances in Software and Hardware for Big Data to Knowledge Discovery (ASH) and 4th International Workshop on Distributed Storage Systems and Coding for Big Data	Dr. Hui Zhang, Dr. Weijia Xu, Dr. Hongfeng Yu, Dr. Bing Zhu	*Regency B*
	W9, The 1st IEEE International Workshop on Big Spatial Data (BSD)	Dr. Chengyang Zhang, Dr. Abdeltawab Hendawi, Dr. Farnoush Banaei-kashani	*Regency C*
	W15, IEEE Workshop on Big Data Metadata and Management (BDMM 2016)	Dr. Alex Kuo, Dr. Yinglong Xia, Dr. Mahmoud Daneshmand, Dr. Chonggang Wang	*Regency D*
	W21, The 3rd Workshop on Pattern Mining and Application of Big Data (BigPMA 2016)	Dr. Yi-Cheng Chen, Dr. Jiun-Long Huang	*Columbia A*

Monday, 5-December - continued			
Time	**Sessions/Workshops**	**Session Chair**	**Location**
Sessions 8:00 – 12:00 pm	**W5**, Real-time and Stream Analytics in Big Data Workshop	Nam-Luc TRAN, Sabri SKHIRI and Thomas PEEL	*Columbia B*
	W10, IEEE Workshop on Big Data and Machine Learning in Telecom (BMLIT)	Dr. Jin Yang, Dr. Hui Zang,Dr. Li Liu	*Columbia C*
	W13, Big Data Challenges, Research, and Technologies in the Earth and Planetary Sciences	Dr. Tom Narock, Dr. Dan Crichton	*Congressional C*
	W16, Workshop on Big Data in Smart Grids	Dr. Wilson Rivera	*Concord*
	W19, 3rd International Workshop on High Performance Big Graph Data Management, Analysis, and Mining (BigGraphs 2016)	Dr. Kamesh Madduri, Dr. Mohammad Al Hasan, Dr. Nesreen Ahmed	*Columbia Foyer*
	W20, 3rd Big Data Analytic Technology for Bioinformatics and Health Informatics (KDDBHI) Workshop	Dr.Donghui Wu, Dr. Xin Deng, Ph.D.	*Lexington*
	W27, Textual Customer Feedback Mining and Transfer Learning Workshop	Dr. Xin Deng, Dr. Ming Shao, Dr. Ross Smith, Dr. Yun Fu	*Congressional B*
12:00 - 1:30 pm	**Lunch (on own)**		
Time	**Sessions/Workshops/Tutorials**	**Session Chair**	**Location**
Tutorial 1:30 – 6:00 pm	**Tutorial 1**: Large Scale Text Mining – Techniques and Applications (1:30-3:30 pm)	Prof. Ronen Feldman, Prof. Ron Bekkerman	*Columbia B*
	Tutorial 5: Anomalous and Significant Subgraph Detection in Attributed Networks (4:00 -6:00 pm)	Feng Chen, Petko Bogdanov, Daniel B. Neill, Ambuj K. Singh	*Columbia B*
Sessions 1:30 – 6:30 pm	**W1**, Big Data for Cloud Operations Management: Problems, Approaches, Tools, and Best Practices Workshop	Dr. Ata Turk	*Columbia C*
	W4 First Workshop on Open Science in Big Data (OSBD)	Dr. Shannon Quinn	*Concord*
	W7, Methods to Manage Heterogeneous Big Data and Polystore Databases Workshop	Dr. Vijay Gadepally	*Lexington*
	W12, 2nd International workshop on Methodologies to Improve Big Data Project	Dr. Jeffrey Saltz	*Bunker Hill*
	W17, 3rd Solar & Stellar Astronomy Big Data (SABiD) - Workshop on Management, Search and Mining of Massive Repositories of Solar and Stellar Astronomy Data	Dr. Rafal A. Angryk, Dr. Piet C. Martens, Dr. Russel J. White	*Congressional C*
	W22, Advances in High Dimensional (AdHD) Big Data	Dr. Sotiris Tasoulis, Dr. Liang Wang	*Congressional D*
	W26, Ieee Workshop On Big Data Analytics In Manufacturing And Supply Chains	Dr. Zhang Nengsheng Allan, Ms. Mika Kawai	*Columbia Foyer*
	W28, Big Data and Natural Language Processing (Big-NLP-2016) Workshop	Dr. Paul Rayson	*Congressional B*

Tuesday, 6-December			
7:20-6:00 pm Location:	**Registration** *Regency Foyer B Wall*		
Time	**Sessions**	**Session Chair**	**Location**
8:30-08:45	Opening and Welcome	Sudarsan Rachuri, Lyle Ungar, Philip S. Yu, James Joshi, Ling Liu, George Karypis	*Regency ABCD*
8:45-09:45	**Keynote Session 1**: Database Decay and How to Avoid It *Dr. Michael Stonebraker, Paradigm4/MIT, USA*	Ling Liu	*Regency ABCD*
9:45-10:45	**Keynote Session 2**: Leveraging High Performance Computing to Drive Advanced Manufacturing R&D at the US Department of Energy *Dr. Mark Johnson, Director, Advanced Manufacturing Office, U.S. Department of Energy*	Sudarsan Rachuri	*Regency ABCD*
8:30-10:45	**Special Session**: Intelligent Data Mining	Dr. Uraz Yavanoglu	*Concord*
10:45 – 11:05 am Location:	**Coffee Break** *Regency Foyer* **Poster Session (Set up only)** *Regency Foyer and Hall of Battles*		
11:05 am -12:45 pm	**Sessions**	**Session Chair**	**Location**
	L1 Cloud/Stream Computing for Big Data	Latifur Khan, UT Dallas	*Regency ABC*
	L2 Link and Graph Mining I	John A Miller, University of Georgia	*Columbia A*
	L3 Visual Analytics and Mobility	Athanasios V. Vasilakos, Lulea University of Technology	*Columbia B*
	L4 Big Data in Healthcare	Aki-Hiro Sato, Kyoto University	*Regency D*
	I&G-Regular 1: Big Data Analytics	Shakti Awaghad	*Columbia Foyer*
	Manufacturing Symposium	Dr. Sudarsan Rachuri, Tina Lee, Dr. Ronay Ak, Dr. Anantha Narayanan, Dr. Soundar Srinivasan , Dr. Rumi Ghosh , Dr. Steve Eglash	*Columbia C*
	Special Session: Intelligent Data Mining	Dr. Uraz Yavanoglu	*Concord*
	W6 Application of Big Data for computational social science	Dr. Akira Ishii,Dr. Fujio Toriumi, Dr. Yasuko Kawahata	*Lexington*
12:45 – 2:00 pm Location:	**Lunch (provided by conference)** *Regency ABC* **Poster Session Sets Up and Displays** *Regency Foyer and Hall of Battles*		

2:00 – 4:05 pm	**Sessions**	**Session Chair**	**Location**
	L5 High Performance Platforms for Big Data	John A Miller, University of Georgia	*Regency ABC*
	L6 Spatiotemporal and Stream Data Management	Jianwu Wang, UMBC	*Columbia A*
	L7 Big Data Processing/Mining	Athanasios V. Vasilakos, Lulea University of Technology	*Columbia B*
	L8 Big Data Applications I	Aki-Hiro Sato, Kyoto University	*Regency D*
	I&G-Regular 2: Big Data Applications (1)	Yinglong Xia	*Columbia Foyer*
	Manufacturing Symposium	Dr. Sudarsan Rachuri, Tina Lee, Dr. Ronay Ak, Dr. Anantha Narayanan, Dr. Soundar Srinivasan , Dr. Rumi Ghosh , Dr. Steve Eglash	*Columbia C*
	Special Session: Intelligent Data Mining	Dr. Uraz Yavanoglu	*Concord*
	W6 Application of Big Data for computational social science	Dr. Akira Ishii,Dr. Fujio Toriumi, Dr. Yasuko Kawahata	*Lexington*
4:05 – 4:25 pm Location:	**Coffee Break** *Regency Foyer* **Poster Session Sets Up and Displays** *Regency Foyer and Hall of Battles*		
4:25 -6:25 pm	**Sessions**	**Session Chair**	**Location**
	Panel 1 (Moderator: Ling Liu)	Moderator: Ling Liu, Georgia Institute of Technology	*Regency ABC*
	S1 Visualization, Multimedia, & Crowdsourcing	Haiying Shen, University of Virginia	*Columbia A*
	S2 Computational Models, and Social Media Recommendation	Panagiotis Liakos, University of Athens	*Columbia B*
	S3 Energy-Efficiency and Data Quality/Processing	Yongluan Zhou. University of Southern Denmark	*Regency D*
	I&G –Short 1: Big Data Algorithms & Systems	Dr. Michael E. Sharp	*Columbia Foyer*
	Manufacturing Symposium	Dr. Sudarsan Rachuri, Tina Lee, Dr. Ronay Ak, Dr. Anantha Narayanan, Dr. Soundar Srinivasan , Dr. Rumi Ghosh , Dr. Steve Eglash	*Columbia C*
	Special Session: Intelligent Data Mining	Dr. Uraz Yavanoglu	*Concord*
	W6 Application of Big Data for computational social science	Dr. Akira Ishii,Dr. Fujio Toriumi, Dr. Yasuko Kawahata	*Lexington*

Wednesday, 7-December			
7:30-6:00 pm Location:	**Registration** *Regency Foyer B Wall*		
8:30 – 8:45	**Opening remarks / Announcements**		
Time	**Sessions**	**Session Chair**	**Location**
8:45 -9:45 am	**Keynote Speech 3**: Harnessing the Data Revolution: A Perspective from the National Science Foundation *Dr. Chaitanya Baru, National Science Foundation*	Ling Liu	*Regency ABCD*
9:45 -10:45 am	**Keynote Speech 4**: On the Power of Big Data: Mining Structures from Massive, Unstructured Text Data *Prof. Jiawei Han, Abel Bliss Professor, University of Illinois at Urbana-Champaign, USA*	James Joshi	*Regency ABCD*
8:00-10:45am	**W25**, 3rd International Workshop on Privacy and Security of Big Data (PSBD 2016)	Dr. Alfredo Cuzzocrea	*Bunker Hill*
10:45 - 11:05am Location:	**Coffee Break** *Regency Foyer* **Poster Session Displays** *Regency Foyer and Hall of Battles*		
11:05- 12:45 pm	**Sessions**	**Session Chair**	**Location**
	L9 Link and Graph Mining II	Jun (Luke) Huan, University of Kansas	*Regency ABC*
	L10 Social Networks/Media	Weijia Xu, UT Austin	*Columbia A*
	L11 Big Data Applications II	Aki-Hiro Sato, Kyoto University	*Columbia B*
	L12 Stream Data Mining /Cloud - Big Velocity Data	Jianwu Wang, UMBC	*Regency D*
	I&G-Regular 3: Big Data Platforms & Frameworks	Pavan Kapanipathil	*Columbia Foyer*
	Manufacturing Symposium	Dr. Sudarsan Rachuri, Tina Lee, Dr. Ronay Ak, Dr. Anantha Narayanan, Dr. Soundar Srinivasan , Dr. Rumi Ghosh , Dr. Steve Eglash	*Columbia C*
	W11 4th Workshop on Scalable Cloud Data Management Workshop (SCDM)	Mr. Felix Gessert	*Concord*
	W25, 3rd International Workshop on Privacy and Security of Big Data (PSBD 2016)	Dr. Alfredo Cuzzocrea	*Bunker Hill*
12:45 - 2:00 pm Location:	**Lunch (provided by Conference)** *Regency ABC* **Poster Session Displays** Regency Foyer and Hall of Battles		
Time	**Sessions**	**Session Chair**	**Location**
2:00 – 4:05 pm	**L13** Big Data Analytics and Security/Privacy I	Zhiyuan Chen, UMBC	*Regency ABC*
	L14 Architecture/Systems and Big Data Analytics	Weijia Xu, UT Austin	*Columbia A*
	L15 Data Management & Applications	Aki-Hiro Sato, Kyoto University	*Columbia Foyer*
	L16 Algorithms and Systems for Big Data Search	Athanasios V. Vasilakos, Lulea University of Technology	*Regency D*
	I&G Panel Session (2:00pm~3:00pm) Big Data Regional Innovation Hubs: Accelerating the Big	Dr. Lea Shanley	*Columbia B*

	Data Innovation Ecosystem		
	I&G-short2 (3:00pm~4:05pm) Massive Processing & Experience	Dr. William Z. Bernstein	*Columbia B*
	Manufacturing Symposium	Dr. Sudarsan Rachuri, Tina Lee, Dr. Ronay Ak, Dr. Anantha Narayanan, Dr. Soundar Srinivasan , Dr. Rumi Ghosh , Dr. Steve Eglash	*Columbia C*
	W11 4th Workshop on Scalable Cloud Data Management Workshop (SCDM)	Mr. Felix Gessert	*Concord*
	W25, 3rd International Workshop on Privacy and Security of Big Data (PSBD 2016)	Dr. Alfredo Cuzzocrea	*Bunker Hill*
4:05 – 4:25 pm Location	**Coffee Break** *Regency Foyer* **Poster Session Displays** *Regency Foyer and Hall of Battles*		
Time	**Sessions**	**Session Chair**	**Location**
4:25- 6:25 pm	**Panel 2** (Moderator: James Joshi)	Moderator: Eui-Hong (Sam) Han, The Washington Post	*Columbia B*
	S4 Link and Graph Mining III	Jun (Luke) Huan, University of Kansas	*Columbia A*
	S5 Big Data Analytics and Security/Privacy II	Zhiyuan Chen, UMBC	*Bunker Hill*
	S6 Algorithms and Systems for Big Data II	Dirk Van den Poel, Ghent University	*Regency D*
	I&G-regular4: Big Data Applications (2)	Dr. Lijun Qian	*Columbia Foyer*
	Manufacturing Symposium	Dr. Sudarsan Rachuri, Tina Lee, Dr. Ronay Ak, Dr. Anantha Narayanan, Dr. Soundar Srinivasan , Dr. Rumi Ghosh , Dr. Steve Eglash	*Columbia C*
	W11 4th Workshop on Scalable Cloud Data Management Workshop (SCDM)	Mr. Felix Gessert	*Concord*
7:00 – 9:00 pm Location	**Banquet (Ticket required)** *Regency ABC* 1. *Best Paper Award,* PC Co-chairs 2. *Best Application Paper Award*		

Thursday, 8-December			
07:30-6:00pm Location:	**Registration** *Regency Foyer B Wall*		
Time	**Session**	**Session Chair**	**Location**
8:30 – 9:45 am	**Opening Remarks / Announcements**		
8:45 - 09:45 am	**Keynote Speech 5:** Big Data Security and Privacy *Prof. Elisa Bertino, Purdue University, USA*	James Joshi	*Regency ABC*
9:45 - 10:45 am	**Keynote Speech 6:** Cognitive Computing: From breakthroughs in the lab to applications on the field *Dr. Guruduth S. Banavar, Vice President and Chief Science Officer, Cognitive Computing, IBM*	James Joshi	*Regency ABC*
8:00-10:45am	**Granular Computing Special Session**: Data Science and Computing	T.Y. Lin	*Columbia Foyer*
10:45 - 11:05 am Location:	**Coffee Break** *Regency Foyer* **Poster Session Displays** *Regency Foyer and Hall of Battles*		
Time	**Sessions/Tutorial/Workshop**	**Session Chair**	**Location**
11:05am – 12:45pm	**L17** Computational Models for BigData I	Seung-Jong Jay Park, Louisiana State University	*Regency A*
	L18 Computational Models for BigData II	Panagiotis Liakos, University of Athens	*Regency B*
	S7 Theoretical Models for Big Data	Alfredo Cuzzocrea University of Trieste	*Regency C*
	S8 Software Systems/Platform for Big Data Computing	Kyong Jin Shim, Singapore Management University	*Regency D*
	Tutorial 2: Trajectory Data Mining (11am-1pm)	Prof. Zhenhui (Jessie) Li, Fei Wu, Prof. Jiawei Han	*Columbia Foyer*
	Tutorial 3: Large Scale Matrix Factorization(11am-1pm)	Fei Wang, Wei Tan	*Concord*
	Tutorial 4: Dynamic Big Data Processing in the Web of Things: Challenges, Opportunities and Success Stories (11am-1pm)	Ljiljana Stojanovic, Nenad Stojanovic	*Lexington*
Whole day workshop 8:30am - 6pm	**W18**, Computational Archival Science: digital records in the age of big data	Dr. Mark Hedges	*Bunker Hill*
12:45- 2:00 pm	**Lunch (provided by conference)** *Regency ABC* **Poster Session Displays** *Regency Foyer and Hall of Battles*		
Time	**Sessions/Workshops**	**Session Chair**	**Location**
2:00 – 4:30	**S9** Cloud/High Performance/Parallel Computing and Big Data	Athanasios N. Nikolakopoulos, University of Patras	*Regency A*
	S10 Big Data Applications III	Kyong Jin Shim, Singapore Management University	*Regency B*
	S11 Big Data Search and Mining in Social Media and Web	Panagiotis Liakos, University of Athens	*Regency C*
	S12 Data Management & Integration	Alfredo Cuzzocrea University of Trieste	*Regency D*
3:30 –4:00 pm	**Coffee Breeak**		

Keynote Lectures

Keynote 1: Database Decay and How to Avoid It

Speaker:
Dr. Michael Stonebraker, Paradigm4/MIT, USA

Abstract:
The traditional wisdom for designing database schemas is to use a design tool (typically based on a UML or ER model) to construct an initial data model for one's data. When one is satisfied with the result, the tool will automatically construct a collection of 3rd normal form relations for the model. Then applications are coded against this relational schema. When business circumstances change (as they do frequently) one should run the tool again to produce a new data model and a new resulting collection of tables. The new schema is populated from the old schema, and the applications are altered to work on the new schema, using relational views whenever possible to ease the migration. In this way, the database remains in 3rd normal form, which represents a "good" schema, as defined by DBMS researchers. "In the wild", schemas often change once a quarter or more often, and the traditional wisdom is to repeat the above exercise for each alteration. In this paper we report that the traditional wisdom appears to be rarely-to-never followed "in the wild" for large, multi-department applications. Instead DBAs appear to attempt to minimize application maintenance (and hence schema changes) instead of maximizing schema quality. This leads to schemas which quickly diverge from ER or UML models and actual database semantics tend to drift farther and farther from 3rd normal form.We term this divergence of reality from 3rd normal form principles database decay. Obviously, this is a very undesirable state of affairs. In this paper we explore the reasons for database decay and tactics to avoid it. These include defensive schemas, defensive application programs and a different model for interacting with a database

Short Bio:

Dr. Stonebraker has been a pioneer of data base research and technology for more than forty years. He was the main architect of the INGRES relational DBMS, and the object-relational DBMS, POSTGRES. These prototypes were developed at the University of California at Berkeley where Stonebraker was a Professor of Computer Science for twenty five years. More recently at M.I.T. he was a co-architect of the Aurora/Borealis stream processing engine, the C-Store column-oriented DBMS, the H-Store transaction processing engine, the SciDB array DBMS, and the Data Tamer data curation system. Presently he serves as Chief Technology Officer of Paradigm4 and Tamr, Inc.Professor Stonebraker was awarded the ACM System Software Award in 1992 for his work on INGRES. Additionally, he was awarded the first annual SIGMOD Innovation award in 1994, and was elected to the National Academy of Engineering in 1997. He was awarded the IEEE John Von Neumann award in 2005 and the 2014 Turing Award, and is presently an Adjunct Professor of Computer Science at M.I.T, where he is co-director of the Intel Science and Technology Center focused on big data.

Keynote 2: Leveraging High Performance Computing to Drive Advanced Manufacturing R&D at the US Department of Energy

Speaker:
Dr. Mark Johnson, Director, Advanced Manufacturing Office, U.S. Department of Energy

Abstract:
Manufacturing is a critical component of the U.S. economy, responsible for 12.5% of GDP, direct employment for over 12 million people, and close to 75% of U.S. exports of goods. The U.S. manufacturing sector, while it produces 17% of the world's manufacturing output, also represents a quarter of the country's energy consumption. On the R&D side, it is responsible for 70% of all private sector R&D performed (in 2010 and 2011) and nearly 60% of patent applications. A number of emerging technologies are driving shifts in traditional manufacturing, in particular the convergence of information and communication technology with the materials and process technologies of manufacturing. Particularly for energy intensive and energy-dependent

industries, harnessing IT to reduce energy usage while simultaneously making companies more competitive is essential to the future of U.S. manufacturing, competitiveness, and productivity.

This talk will review the Advanced Manufacturing Offices work to leverage high performance computing, smart manufacturing approaches for the U.S. clean energy manufacturing sector—through targeted R&D in modeling and simulation and partnerships with industry, academia, technology incubators and other stakeholders.

Short Bio:

Mark Johnson, Ph.D. serves as the Director of the Advanced Manufacturing Office (AMO) in the Office of Energy Efficiency and Renewable Energy (EERE). AMO is focused on creating a fertile innovation environment for advanced manufacturing, enabling vigorous domestic development of new energy-efficient manufacturing processes and materials technologies to reduce the energy intensity and life-cycle energy consumption of manufactured products.

Previously, Mark served as a Program Director in the Advanced Research Projects Agency–Energy (ARPA-E) where he had the longest tenure in that post—from ARPA-E's formation in 2010 to mid-2013. At ARPA-E, Mark led initiatives to advance energy storage and critical materials, as well as projects in small business, advanced semiconductor, novel wind architectures, superconductors and electric machines

He also served as the Industry and Innovation Program Director for the Future Renewable Electric Energy Delivery and Management (FREEDM) Systems Center. This is a National Science Foundation Gen-111 Engineering Research Center targeting the convergence of power electronics, energy storage, renewable resource integration and information technology for electric power systems.

Mark joins EERE on assignment from North Carolina State University, where he is an Associate Professor of Materials Science and Engineering. His research has focused on crystal growth and device fabrication of compound semiconductor materials with electronic and photonic applications. Mark also taught in the Technology, Entrepreneurship and Commercialization program jointly between the NC State Colleges of Management and Engineering. In addition to his academic career, Mark is an entrepreneur and early stage leader in Quantum Epitaxial Designs (now International Quantum Epitaxy), EPI Systems (now Veeco) and Nitronex (now GaAs Labs).

Mark has a bachelor's degree from MIT and a Ph.D., from NC State, both in Materials Science and Engineering.

Keynote 3: Harnessing the Data Revolution: A Perspective from the National Science Foundation

Speaker:
Dr. Chaitanya Baru, NSF, USA

Abstract:
This talk will introduce NSF's vision for moving beyond initial, isolated approaches for data science research, services, and infrastructure, towards a cohesive, federated, national-scale approach to harness the data revolution and transform US science, engineering, and education over the next decade and beyond.

Short Bio:

Chaitan Baru is Senior Advisor for Data Science in the Computer and Information Science and Engineering (CISE) Directorate at the National Science Foundation. He is there on assignment from the San Diego Supercomputer, UC San Diego, where he is Associate Director for Data Initiatives. At NSF, he coordinates the cross-Foundation BIGDATA research program, advises the NSF Big Data Hubs and Spokes program, assists in strategic planning, and participates in interdisciplinary and inter-agency Data Science-related activities. He co-chairs the Big Data Inter-agency Working Group, and is co-author of the US Federal Big Data R&D Strategic Plan (https://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/bigdatardstrategicplan-nitrd_final-051916.pdf), released in May 2016 under the auspices of the Networking and Information Technology R&D (NITRD) group of the National Coordination Office, White House Office of Science

and Technology Policy.

Keynote 4: On the Power of Big Data: Mining Structures from Massive, Unstructured Text Data

Speaker:
Prof. Jiawei Han, Abel Bliss Professor, University of Illinois at Urbana-Champaign, USA

Abstract:
The real-world big data are largely unstructured, interconnected, and in the form of natural language text. One of the grand challenges is to turn such massive unstructured data into structured ones, and then to structured networks and actionable knowledge. We propose a data-intensive text mining approach that requires only distant supervision or minimal supervision but relies on massive data. We show quality phrases can be mined from such massive text data, types can be extracted from massive text data with distant supervision, and relationships among entities can be discovered by meta-path guided network embedding. Finally, we propose a D2N2K (i.e., data-to-network-to-knowledge) paradigm, that is, first turn data into relatively structured information networks, and then mine such text-rich and structure-rich networks to generate useful knowledge. We show such a paradigm represents a promising direction at turning massive text data into structured networks and useful knowlege.

Short Bio:

Jiawei Han is Abel Bliss Professor in the Department of Computer Science, University of Illinois at Urbana-Champaign. He has been researching into data mining, information network analysis, database systems, and data warehousing, with over 700 journal and conference publications. He has chaired or served on many program committees of international conferences, including PC co-chair for KDD, SDM, and ICDM conferences, and Americas Coordinator for VLDB conferences. He also served as the founding Editor-In-Chief of ACM Transactions on Knowledge Discovery from Data and the Director of Information Network Academic Research Center supported by U.S. Army Research Lab, and is the co-Director of KnowEnG, an NIH funded Center of Excellence in Big Data Computing. He is a Fellow of ACM and Fellow of IEEE, and received 2004 ACM SIGKDD Innovations Award, 2005 IEEE Computer Society Technical Achievement Award, 2009 M. Wallace McDowell Award from IEEE Computer Society. His co-authored book "Data M ining: Concepts and Techniques" has been adopted as a textbook popularly worldwide.

Keynote 5: Big Data Security and Privacy

Speaker:
Prof. Elisa Bertino, Purdue University, USA

Abstract:
Technological advances and novel applications, such as sensors, cyber-physical systems, smart mobile devices, cloud systems, data analytics, and social networks, are making possible to capture, and to quickly process and analyze huge amounts of data from which to extract information critical for security-related tasks. In the area of cyber security, such tasks include user authentication, access control, anomaly detection, user monitoring, and protection from insider threat. By analyzing and integrating data collected on the Internet and Web one can identify connections and relationships among individuals that may in turn help with homeland protection. By collecting and mining data concerning user travels and disease outbreaks one can predict disease spreading across geographical areas. And those are just a few examples; there are certainly many other domains where data technologies can play a major role in enhancing security. The use of data for security tasks is however raising major privacy concerns. Collected data, even if anonymized by removing identifiers such as names or social security numbers, when linked with other data may lead to re-identify the individuals to which specific data items are related to. Also, as organizations, such as governmental agencies, often need to collaborate on security tasks, data sets are exchanged across different organizations, resulting in these data sets being available to many different parties. Apart from the use of data for analytics, security tasks such as authentication and access control may require detailed information about users. An example is multi-factor authentication that may require, in addition to a password or a certificate, user biometrics. Recently proposed continuous authentication techniques extend access control system. This information if misused or stolen can lead to privacy breaches.It would then seem that in order to achieve security we must give up privacy. However this may not be necessarily the case. Recent advances in cryptography are making possible to work on encrypted data – for example for performing analytics on encrypted data. However much more needs to be done as the specific data privacy techniques to use heavily depend on the specific use of data and the security tasks at hand.

Also current techniques are not still able to meet the efficiency requirement for use with big data sets. In this talk we will discuss methods and techniques to make this reconciliation possible and identify research directions.

Short Bio:

Elisa Bertino is professor of computer science at Purdue University, and serves as Research Director of the Center for Information and Research in Information Assurance and Security (CERIAS). She is also an adjunct professor of Computer Science & Info tech at RMIT. Prior to joining Purdue in 2004, she was a professor and department head at the Department of Computer Science and Communication of the University of Milan. She has been a visiting researcher at the IBM Research Laboratory (now Almaden) in San Jose, at the Microelectronics and Computer Technology Corporation, at Rutgers University, at Telcordia Technologies. Her recent research focuses on data security and privacy, digital identity management, policy systems, and security for the Internet-of-Things. She is a Fellow of ACM and of IEEE. She received the IEEE Computer Society 2002 Technical Achievement Award, the IEEE Computer Society 2005 Kanai Award, and the ACM SIGSAC 2014 Outstanding Contributions Award. She is currently serving as EiC of IEEE Transactions on Dependable and Secure Computing.

Keynote 6: Cognitive Computing: From breakthroughs in the lab to applications on the field

Speaker:
Dr. Guruduth S. Banavar, Vice President and Chief Science Officer, Cognitive Computing, IBM

Abstract:
In the last decade, the availability of massive amounts of new data, the development of new machine learning technologies, and the availability of scalable computing infrastructure, have given rise to a new class of computing systems. These "Cognitive Systems" learn from data, reason from models, and interact naturally with us, to perform complex tasks better than either humans or machines can do by themselves. These tasks range from answering questions conversationally to extracting knowledge for discovering insights to evaluating options for difficult decisions. These cognitive systems are designed to create new partnerships between people and machines to augment and scale human expertise in every industry, from healthcare to financial services to education. This talk will provide an overview of cognitive computing, the technology breakthroughs that are enabling this trend, and the practical applications of this technology that are transforming every industry.

Short Bio:

Dr. Guru Banavar is vice president and chief science officer for cognitive computing at IBM. He is responsible for advancing the next generation of cognitive technologies and solutions with IBM's global scientific ecosystem, including academia, government agencies and other partners. Most recently, he led the team responsible for creating new AI technologies and systems in the family of IBM Watson, designed to augment human expertise in all industries. Previously, as chief technology officer for IBM's Smarter Cities initiative, Banavar designed and implemented big data and analytics systems to help make cities, such as Rio de Janeiro and New York, more livable and sustainable. Prior to that, he was director of IBM Research in India, where he and his team received a presidential award for improving technology access with the Spoken Web project. Across his career, Banavar and his team have delivered a range of products and solutions for IBM and its clients. He has also served on task forces such as NY Governor Cuomo's commission to improve resilience to natural disasters. He holds more than 25 patents and has published extensively, with his work featured in media outlets around the world.

Conference Paper Presentations

L1: Cloud/Stream Computing for Big Data	
Regular	BigD215 "CCRP: Customized Cooperative Resource Provisioning for High Resource Utilization in Clouds" Jinwei Liu, Haiying Shen, and Husnu Narman
Regular	BigD431 "Towards Resource-Efficient Cloud Systems: Avoiding Over-Provisioning in Demand-Prediction Based Resource Provisioning" Liuhua Chen and Haiying Shen
Regular	BigD437 "A Low-Load Stream Processing Scheme for IoT Environments" Tomoki Yoshihisa and Takahiro Hara
Regular	BigD540 "Matrix Factorizations at Scale: a Comparison of Scientific Data Analytics in Spark and C+MPI Using Three Case Studies" Alex Gittens, Aditya Devarakonda, Evan Racah, Michael Ringenburg, Lisa Gerhardt, Jey Kottalam, Jialin Liu, Kristyn Maschhoff, Shane Canon, Jatin Chhugani, Pramod Sharma, Jiyan Yang, James Demmel, Jim Harrell, Venkat Krishnamurthy, Michael Mahoney, and Mr Prabhat

L2: Link and Graph Mining I	
Regular	BigD553 "Local Graphlet Estimation" Nesreen Ahmed, Ted Willke, and Ryan Rossi
Regular	BigD595 "Parallel Top-k Subgraph Query in Massive Graphs: Computing from the Perspective of Single Vertex" Jianliang Gao, Bo Song, Ping Liu, Weimao Ke, Jianxin Wang, and Xiaohua Hu
Regular	BigD628 "Dynamic Feature Generation and Selection on Heterogeneous Graph for Music Recommendation" Chun Guo and Xiaozhong Liu
Regular	BigD638 "Community Detection with Partially Observable Links and Node Attributes" Xiaokai Wei, Bokai Cao, Weixiang Shao, Chun-Ta Lu, and Philip S. Yu

L3: Visual Analytics and Mobility	
Regular	BigD404 "TelcoFlow: Visual Exploration of Collective Behaviors Based on Telco Data" Yixian Zheng, Wenchao Wu, Haipeng Zeng, Nan Cao, Huamin Qu, Mingxuan Yuan, Jia Zeng, and Lionel M. Ni
Regular	BigD586 "PRIIME: A Generic Framework for Interactive Personalized Interesting Pattern Discovery" Mansurul Bhuiyan and Mohammad Al Hasan
Regular	BigD278 "Predicting Taxi Demand at High Spatial Resolution: Approaching the Limit of Predictability" Kai Zhao, Denis Khryashchev, Juliana Freire, Claudio Silva, and Huy Vo
Regular	BigD346 "Parallel Gathering Discovery over Big Trajectory Data" Yongyi Xian, Yan Liu, and Chuanfei Xu

L4: Big Data in Healthcare	
Regular	BigD237 "Real-time Full Correlation Matrix Analysis of fMRI Data" Yida Wang, Bryn Keller, Mihai Capotă, Michael Anderson, Narayanan Sundaram, Jonathan Cohen, Kai Li, Nicholas Turk-Browne, and Ted Willke
Regular	BigD383 "Enabling Factor Analysis on Thousand-Subject Neuroimaging Datasets" Michael Anderson, Mihai Capotă, Javier Turek, Xia Zhu, Ted Willke, Yida Wang, Po-Hsuan Chen, Jeremy Manning, Peter Ramadge, and Kenneth Norman
Regular	BigD503 "Using Machine Learning to Identify Major Shifts in Human Gut Microbiome Protein Family Abundance in Disease" Mehrdad Yazdani, Bryn Taylor, Justine Debelius, Weizhong Li, Rob Knight, and Larry Smarr
Regular	BigD543 "Network Analysis for Identifying and Characterizing Disease Outbreak Influence from

	Voluminous Epidemiology Data" Naman Shah, Harshil Shah, Matthew Malensek, Sangmi Lee Pallickara, and Shrideep Pallickara

L5: High Performance Platforms for Big Data	
Regular	BigD252 "Spark-GPU: High Performance In-Memory Data Processing with GPUs" Yuan Yuan, Meisam Fathi, Kaibo Wang, Rubao Lee, and Xiaodong Zhang
Regular	BigD363 "Comparing Application Performance on HPC-based Hadoop Platforms with Local Storage and Dedicated Storage" Zhuozhao Li, Haiying Shen, Jeffrey Denton, and Walter Ligon
Regular	BigD538 "Thrill: High-Performance Algorithmic Distributed Batch Data Processing with C++" Timo Bingmann, Michael Axtmann, Emanuel Jöbstl, Sebastian Lamm, Huyen Chau Nguyen, Alexander Noe, Sebastian Schlag, Matthias Stumpp, Tobias Sturm, and Peter Sanders
Regular	BigD607 "High-Performance Design of Apache Spark with RDMA and Its Benefits on Various Workloads" Xiaoyi Lu, Dipti Shankar, Shashank Gugnani, and Dhabaleswar K. Panda
Regular	BigD597 "Efficient Data Access Strategies for Hadoop and Spark on HPC Cluster with Heterogeneous Storage" Nusrat Islam, Md Wasi-ur- Rahman, Xiaoyi Lu, and Dhabaleswar K. Panda

L6: Spatiotemporal and Stream Data Management	
Regular	BigD426 "Sampling based Distributed Kernel Mean Matching using Spark" Ahsanul Haque, Zhuoyi Wang, Swarup Chandra, Latifur Khan, and Charu Aggarwal
Regular	BigD515 "Handling Uncertainty in Trajectories of Moving Objects in Unconstrained Outdoor Spaces" Eleazar Leal, Le Gruenwald, and Jianting Zhang
Regular	BigD526 "Clockwise Compression for Trajectory Data under Road Network Constraints" Yudian Ji, Yuda Zang, Wuman Luo, Xibo Zhou, Ye Ding, and Lionel M. Ni
Regular	BigD569 "In Pursuit of Outliers in Multi-dimensional Data Streams" Shiblee Sadik, Le Gruenwald, and Eleazar Leal
Regular	BigD626 "WISDOM: Weighted Incremental Spatio-Temporal Multi-Task Learning via Tensor Decomposition" Jianpeng Xu, Jiayu Zhou, Pang-Ning Tan, Xi Liu, and Lifeng Luo

L7: Big Data Processing/Mining	
Regular	BigD384 "Sentence-level Extraction of Complementary Entities using Large Unlabeled Product Reviews" HU XU, Sihong Xie, Lei Shu, and Philip S. Yu
Regular	BigD507 "Harnessing Relationships for Domain-specific Subgraph Extraction: A Recommendation Use Case" Sarasi Lalithsena, Pavan Kapanipathi, and Amit Sheth
Regular	BigD469 "Online Inference for Time-varying Temporal Dependency Discovery from Time Series" Chunqiu Zeng, Qing Wang, Wentao Wang, Tao Li, and Larisa Shwartz
Regular	BigD519 "Embedding Feature Selection for Large-scale Hierarchical Classification" Azad Naik and Huzefa Rangwala
Regular	BigD283 "Multi-step Threshold Algorithm for Efficient Feature-based Query Processing in Large-scale Multimedia Databases" Christian Beecks and Alexander Graß

L8: Big Data Applications I	
Regular	BigD371 "Shooting a Moving Target: Motion-Prediction-Based Transmission for 360-Degree Videos" Yanan Bao, Huasen Wu, Tianxiao Zhang, Albara Ramli, and Xin Liu
Regular	BigD471 "Buyer Targeting Optimization: A Unified Customer Segmentation Perspective" Jingyuan Yang, Chuanren Liu, Mingfei Teng, Hui Xiong, and March Liao

Regular	BigD496 "Learning Large-scale Plantation Mapping from Imperfect Annotators" Xiaowei Jia, Ankush Khandelwal, James Gerber, Kimberly Carlson, Paul West, and Vipin Kumar
Regular	BigD517 "When Remote Sensing Data meet Ubiquitous Urban Data: Fine-Grained Air Quality Inference" Yanan Xu and Yanmin Zhu
Regular	BigD623 "Automated IT System Failure Prediction: A Deep Learning Approach" Ke Zhang, Jianwu Xu, Martin Renqiang Min, Guofei Jiang, Konstantinos Pelechrinis, and Hui Zhang

L9: Link and Graph Mining II	
Regular	BigD253 "HEER: Heterogeneous Graph Embedding for Emerging Relation Detection from News" Jingyuan Zhang, Chun-Ta Lu, Mianwei Zhou, Sihong Xie, Yi Chang, and Philip S. Yu
Regular	BigD328 "Efficient Triangle Listing for Billion-Scale Graphs" Hao Zhang, Yuanyuan Zhu, Lu Qin, Hong Cheng, and Jeffrey Xu Yu
Regular	BigD449 "Scalable Link Community Detection: A Local Dispersion-aware Approach" Panagiotis Liakos, Alexandros Ntoulas, and Alex Delis
Regular	BigD541 "Random Surfing on Multipartite Graphs" Athanasios N. Nikolakopoulos, Antonia Korba, and John D. Garofalakis

L10: Social Networks/Media	
Regular	BigD372 "Online Social Network Evolution: Revisiting the Twitter Graph" Hariton Efstathiades, Demetris Antoniades, George Pallis, Marios Dikaiakos, Zoltán Szlávik, and Robert-Jan Sips
Regular	BigD511 "Towards Unsupervised Home Location Inference from Online Social Media" Chao Huang and Dong Wang
Regular	BigD365 "Labeling Actors in Multi-view Social Networks by Integrating Information From Within and Across Multiple Views" Ngot Bui, Thanh Le, and Vasant Honavar
Regular	BigD221 "DistSD: Distance-based Social Discovery with Personalized Posterior Screening" Xiao Pan, Jiawei Zhang, Fengjiao Wang, and Philip S. Yu

L11: Big Data Applications II	
Regular	BigD377 "Scalable genomics: from raw data to aligned reads on Apache YARN" Francesco Versaci, Luca Pireddu, and Gianluigi Zanetti
Regular	BigD455 "Lazer: Distributed Memory-Efficient Assembly of Large-Scale Genomes" Sayan Goswami, Arghya Kusum Das, Richard Platania, Kisung Lee, and Seung-Jong Park
Regular	BigD416 "Leveraging Multi-Granularity Energy Data for Accurate Energy Demand Forecast in Smart Grids" Zhichuan Huang and Ting Zhu
Regular	BigD548 "Ad Allocation with Secondary Metrics" Darja Krushevskaja, S. Muthukrishnan, and Willam Simpson

L12: Stream Data Mining /Cloud - Big Velocity Data	
Regular	BigD355: "An Active Learning Method for Data Streams with Concept Drift" Cheong Hee Park and Youngsoon Kang
Regular	BigD379: "DeltaSherlock: Identifying Changes in the Cloud" Ata Turk, Hao Chen, Anthony Byrne, John Knollmeyer, Sastry Duri, Canturk Isci, and Ayse Coskun
Regular	BigD571: "Distributed and Parallel High Utility Sequential Pattern Mining" Morteza Zihayat, Zane Zhenhua Hu, Aijun An, and Yonggang Hu
Regular	BigD582: "Interpretable and Effective Opinion Spam Detection via Temporal Patterns Mining across Websites" Yuan Yuan, Sihong Xie, Chun-Ta Lu, Philip S. Yu, and Jie Tang

L13: Big Data Analytics and Security/Privacy	
Regular	BigD539 "H2O: A Hybrid and Hierarchical Outlier Detection Method for Large Scale Data Protection" Quan Zhang, Mu Qiao, Ramani Routray, and Weisong Shi
Regular	BigD206 "On the Feasibility of an Embedded Machine Learning Processor for Intrusion Detection" Ricardo Calix and Rajesh Sankaran
Regular	BigD250 "Android Malware Development on Public Malware Scanning Platforms: A Large-scale Data-driven Study" Heqing Huang, Cong Zheng, Junyuan Zeng, Wu Zhou, Sencun Zhu, Peng Liu, Suresh Chari, and Ce Zhang
Regular	BigD439 "Improving the Utility in Differential Private Histogram Publishing: Theoretical Study and Practice" Hui Li, Jiangtao Cui, Xiaobin Lin, and Jianfeng Ma
Regular	BigD451 "Semantic Approach to Automating Management of Big Data Privacy Policies" Karuna Joshi, Aditi Gupta, Sudip Mittal, Claudia Pearce, Anupam Joshi, and Tim Finin

L14: Architecture/Systems and Big Data Analytics	
Regular	BigD335 "ARGO: Architecture-Aware Graph Partitioning" Angen Zheng, Alexandros Labrinidis, Panos Chrysanthis, and Jack Lange
Regular	BigD521 "YinMem: a distributed parallel indexed in-memory computation system for large scale data analytics" Yin Huang, Yelena Yesha, Milton Halem, Yaacov Yesha, and Shujia Zhou
Regular	BigD486 "Mix 'n Match Multi-Engine Analytics" Katerina Doka, Nikolaos Papailiou, Victor Giannakouris, Dimitrios Tsoumakos, and Nectarios Koziris
Regular	BigD448 "Predicting Statistics of Asynchronous SGD Parameters for a Large-Scale Distributed Deep Learning System on GPU Supercomputers" Yosuke Oyama, Akihiro Nomura, Ikuro Sato, Hiroki Nishimura, Yukimasa Tamatsu, and Satoshi Matsuoka
Regular	BigD649 "Adaptive Neuron Apoptosis for Accelerating Deep Learning on Large Scale Systems" Charles Siegel, Jeff Daily, and Abhinav Vishnu

L15: Data Management & Applications	
Regular	BigD456: "Materialized View Selection in Feed Following Systems" Kaiji Chen and Yongluan Zhou
Regular	BigD532: "Accelerating Large-scale Unstructured Mesh Queries" Cuong Nguyen and Rhodes Philip
Regular	BigD367: "MuSQLE: Distributed SQL Query Execution Over Multiple Engine Environments" Victor Giannakouris, Nikolaos Papailiou, Dimitrios Tsoumakos, and Nectarios Koziris
Regular	BigD525: "Improved Methods for Static Index Pruning" Wei Jiang, Juan Rodriguez, and Torsten Suel
Regular	BigD315: "Pairwise Topic Model and its Application to Topic Transition and Evolution," Xiaoli Song and Xiaohua Hu

L16: Algorithms and Systems for Big Data Search	
Regular	BigD260 "Outlier Detection via Sampling Ensemble" Hongfu Liu, Yuchao Zhang, Bo Deng, and Yun Fu
Regular	BigD375 "REQUEST: An Interactive Big Data Exploration Framework" Xiaoyu Ge, Yanbing Xue, Zhipeng Luo, Mohamed Sharaf, and Panos Chrysanthis
Regular	BigD473 "An Adaptive Information-Theoretic Approach for Identifying Temporal Correlations in Big Data Sets" Nguyen Ho, Huy Vo, and Mai Vu

Regular	BigD564 "How Good are Word Embeddings? Automatically Explaining Similarity of Terms" Yating Zhang, Adam Jatowt, and Katsumi Tanaka
Regular	BigD621 "Parallel Computation of k-Nearest Neighbor Joins Using MapReduce" Wooyeol Kim, Younghoon Kim, and Kyuseok Shim

L17: Computational Models for BigData I	
Regular	BigD227 "Cache-oblivious Loops Based on a Novel Space-filling Curve" Christian B&#65533;hm, Martin Perdacher, and Claudia Plant
Regular	BigD318 "A Fast Structured Regression for Large Networks" Fang Zhou, Mohamed Ghalwash, and Zoran Obradovic
Regular	BigD434 "Exact Structure Learning of Bayesian Networks by Optimal Path Extension" Subhadeep Karan and Jaroslaw Zola
Regular	BigD576 "Consensus Optimization with Delayed and Stochastic Gradients on Decentralized Networks" Benjamin Sirb and Xiaojing Ye

L18: Computational Models for BigData II	
Regular	BigD598 "DD-RTREE: A dynamic distributed data structure for efficient data distribution among cluster nodes for spatial data mining algorithms" Poonam Goyal, Jagat Sesh Challa, Nikhil S, Aditya Mangla, Sundar S Balasubramaniam, and Navneet Goyal
Regular	BigD614 "A Meta-graph Approach to Analyze Subgraph-centric Distributed Programming Models" Ravikant Dindokar, Neel Choudhury, and Yogesh Simmhan
Regular	BigD536 "Datalography: Scaling Datalog Graph Analytics on Graph Processing Systems" Walaa Eldin Moustafa, Vicky Papavasileiou, Ken Yocum, and Alin Deutsch

S 1: Visualization, Multimedia, & Crowdsourcing	
Short	BigD300 "Big data on a few pixels" Uwe Jugel, Zbigniew Jerzak, and Volker Markl
Short	BigD226 "A Strategic Approach for Visualizing the Value of Big data (SAVV-BIGD) Framework" Mike Lakoju, Alan Serrano-Rico, and Mark Lycett
Short	BigD299 "Inferring Restaurant Styles by Mining Crowd Sourced Photos from User-Review Websites" Haofu Liao, Yuncheng Li, Tianran Hu, and Jiebo Luo
Short	BigD461 "Efficient Large Scale Near-Duplicate Video Detection Base on Spark" Jinna lv, Bin Wu, Shuai Yang, and Bingjing Jia
Short	BigD506 "Shape Matching using Skeleton Context for Automated Bow Echo Detection" Mohammad Mahdi Kamani, Farshid Farhat, Stephen Wistar, and James Z. Wang
Short	BigD254 "Object Identification with Pay-As-You-Go Crowdsourcing" Ting Wu, Chen Zhang, Lei Chen, Pan Hui, and Siyuan Liu

S 2: Computational Models, and Social Media Recommendation	
Short	BigD487 "Computing Triangle and Open-Wedge Heavy-Hitters in Large Networks" A. Pavan, P. Quint, S. Scott, N. V. Vinodchandran, and J. Smith
Short	BigD643 "Compressed Learning for Time Series Classification" Yuh-Jye Lee, Hsing-Kuo Pao, Shueh-Han Shih, Jing-Yao Lin, and Xin-Rong Chen
Short	BigD319 "Incremental Learning for Matrix Factorization in Recommender Systems" Tong Yu, Ole Mengshoel, Alvin Jude, Eugen Feller, Julien Forgeat, and Nimish Radia
Short	BigD320 "Expenditure Aware Rating Prediction for Recommendation" Chuan Shi, Bowei He, Menghao Zhang, Fuzheng Zhuang, and Philip S. Yu
Short	BigD411 "Context-Aware Point of Interest Recommendation using Tensor Factorization"

	Stathis Maroulis, Ioannis Boutsis, and Vana Kalogeraki
Short	BigD550 "On Robust Truth Discovery in Sparse Social Media Sensing" Daniel Zhang, Rungang Han, and Dong Wang

S 3: Energy-Efficiency and Data Quality/Processing	
Short	BigD357 "Evaluating the Impacts of Code-Level Performance Tunings on Power Efficiency" Satoshi Imamura, Keitaro Oka, Yuichiro Yasui, Yuichi Inadomi, Katsuki Fujisawa, Toshio Endo, Koji Ueno, Keiichiro Fukazawa, Nozomi Hata, Yuta Kakibuka, Koji Inoue, and Takatsugu Ono
Short	BigD537 "Requirements on and Antecedents of Big Data Quality: An Empirical Examination to Improve Big Data Quality in Financial Service Organizations" Adiska Fardani Haryadi, Marijn Janssen, Joris Hulstijn, Haiko Voort, and Agung Wahyudi
Short	BigD445 "Power efficient big data analytics algorithms through low-level operations" Gheorghi Guzun, Josiah McClurg, Guadalupe Canahuate, and Raghuraman Mudumbai
Short	BigD334 "Real Time Processing of Streaming and Static Information" Christoforos Svingos, Theofilos Mailis, Herald Kllapi, Lefteris Stamatogiannakis, Yannis Kotidis, and Yannis Ioannidis
Short	BigD406 "Efficient Processing of Top-k Joins in MapReduce" Mei Saouk, Christos Doulkeridis, Akrivi Vlachou, and Kjetil Noervaag
Short	BigD398 "A Comparison of General-Purpose Distributed Systems for Data Processing" Jinfeng Li, James Cheng, Yunjian Zhao,, Fan Yang, Yuzhen Huang, and Haipeng Chen

S 4: Link and Graph Mining III	
Short	BigD348 "Persistent Cascades: Measuring Fundamental Communication Structure in a Social Network" Steven Morse, Marta Gonzalez, and Natasha Markuzon
Short	BigD408 "Streaming Tensor Summarization" Ioanna Tsalouchidou, Gianmarco De Francisci Morales, Francesco Bonchi, and Ricardo Baeza-Yates
Short	BigD428 "Improving Efficiency of Maximizing Spread in the Flow Authority Model for Large Sparse Networks" Philip Chan and Ebad Ahmadzadeh
Short	BigD446 "Efficient Breadth-First Search on Massively Parallel and Distributed Memory Machines" Koji Ueno, Toyotaro Suzumura, Naoya Maruyama, Katsuki Fujisawa, and Satoshi Matsuoka
Short	BigD454 "Effective and Efficient Graph Augmentation in Large Graphs" Ioanna Filippidou and Yannis Kotidis
Short	BigD646 "Summarizing Large Graphs by Means of Pseudo-Boolean Constraints" Said JABBOUR, Nizar Mhadhbi, Abdesattar Mhadhbi, Badran RAddaoui, and Lakhdar Sais

S 5: Big Data Analytics and Security/Privacy II	
Short	BigD256 "Local Subspace-Based Outlier Detection using Global Neighbourhoods" Bas van Stein, Matthijs van Leeuwen, and Thomas Bäck
Short	BigD475 "Scalable Attack Propagation Model and Algorithms for Honeypot Systems" Ariel Bar, Bracha Shapira, Lior Rokach, and Moshe Unger
Short	BigD302 "Protecting the Location Privacy of Mobile Social Media Users" Shuo Wang, Richard Sinnott, and Surya Nepal
Short	BigD529 "Cloud Kotta: Enabling Secure and Scalable Data Analytics in the Cloud" Yadu Babuji, Kyle Chard, Aaron Gerow, and Eamon Duede
Short	BigD413 "Pick Your Choice in HBase: Security or Performance" Frank Pallas, Johannes Günther, and David Bermbach
Short	BigD338 "Sampling Labelled Profile Data for Identity Resolution" Matthew Edwards, Stephen Wattam, Paul Rayson, and Awais Rashid

S 6: Algorithms and Systems for Big Data II	
Short	BigD333 "Mutiple Submodels Parallel Support Vector Machine on Spark" Chang Liu and Bin Wu
Short	BigD419 "Addressing the Big-Earth-Data Variety Challenge with the Hierarchical Triangular Mesh" Michael Rilee, Kwo-Sen Kuo, Thomas Clune, Amidu Oloso, Paul Brown, and Hongfeng Yu
Short	BigD477 "Three-Hop Distance Estimation in Social Graphs" Pascal Welke, Alexander Markowetz, Torsten Suel, and Maria Christoforaki
Short	BigD514 "Parallel clustering method for Non-Disjoint Partitioning of Large-Scale Data based on Spark Framework" Abir Zayani, Chiheb-Eddine Ben N'Cir, and Nadia Essoussi
Short	BigD601 "Mini-Apps for High Performance Scientific Data Analysis" Sreenivas Sukumar, Michael Matheson, Ramakrishnan Kannan, and Seung-Hwan Lim
Short	BigD373 "Transfer Learning Algorithms for Autonomous Reconfiguration of Wearable Systems" Ramyar Saeedi, Hassan Ghasemzadeh, and Assefaw Gebremedhin

S 7: Theoretical Models for Big Data	
Short	BigD424 "A Theoretical Model for n-gram Distribution in Big Data Corpora" Joaquim Silva, Carlos Goncalves, and Jose Cunha
Short	BigD289 "The self-avoiding walk-jump (SAWJ) algorithm for finding maximum degree nodes in large graphs" Jonathan Stokes and Steven Weber
Short	BigD290 "Efficient multiple scale kernel classifiers" Rocco Langone and Johan A. K. Suykens
Short	BigD316 "Semantic Pattern Mining for Text Mining" Xiaoli Song and Xiaohua Hu
Short	BigD331 "Detecting Gradual Changes from Data Stream Using MDL-Change Statistics" Kenji Yamanishi and Kohei Miyaguchi

S 8: Software Systems/Platform for Big Data Computing	
Short	BigD491 "Big Data Framework Interference In Restricted Private Cloud Settings" Stratos Dimopoulos, Chandra Krintz, and Rich Wolski
Short	BigD236 "Massive Parallelism for Non-linear and Non-stationary Data Analysis with GPGPU" Chun-Chich Chen, Chih-Ya Shen, and Ming-Syan Chen
Short	BigD304 "Performance Evaluation of Big Data Frameworks for Large-Scale Data Analytics" Jorge Veiga, Roberto R. Expósito, Xoán C. Pardo, Guillermo L. Taboada, and Juan Touriño
Short	BigD388 "Adapting to Data Sparsity for Efficient Parallel PARAFAC Tensor Decomposition in Hadoop" Kareem Aggour and Bulent Yener
Short	BigD418 "I'll Take That to Go: Big Data Bags and Minimal Identifiers for Exchange of Large, Complex Data" Kyle Chard, Mike D'Arcy, Ben Heavner, Ian Foster, Carl Kesselman, Ravi Madduri, Alexis Rodriguez, Stian Soiland-Reyes, Carole Goble, Eric Deutsch, Ivo Dinov, Ivo Dinov, Kristi Clark, Nathan Price, and Arthur Toga

S 9: Cloud/High Performance/Parallel Computing and Big Data	
Short	BigD247 "HPTA: High-Performance Text Analytics" Hans Vandierendonck, Karen Murphy, Mahwish Arif, and Dimitrios Nikolopoulos
Short	BigD415 "Entity Resolution Acceleration using Micron's Automata Processor" Chunkun Bo, Ke Wang, Jefferey Fox, and Kevin Skadron
Short	BigD440 "Java Thread and Process Performance for Parallel Machine Learning on Multicore HPC Clusters" Saliya Ekanayake, Supun Kamburugamuve, Pulasthi Wickramasinghe, and Geoffrey Fox

Short	BigD534 "I/O Chunking and Latency Hiding Approach for Out-of-core Sorting Acceleration using GPU and Flash NVM" Hitoshi Sato, Ryo Mizote, Satoshi Matsuoka, and Hirotaka Ogawa
Short	BigD561 "Evaluating the Impact of Data Placement to Spark and SciDB with an Earth Science Use Case" Khoa Doan, Amidu Oloso, Kwo-Sen Kuo, Thomas Clune, and Hongfeng Yu
Short	BigD588 "Boldio: A Hybrid and Resilient Burst-Buffer Over Lustre for Accelerating Big Data I/O" Dipti Shankar, Xiaoyi Lu, and Dhabaleswar K. Panda
Short	BigD644 "Kaleido: Enabling Efficient Scientific Data Processing on Big-Data Systems" Saman Biookaghazadeh, Yiqi Xu, Shujia Zhou, and Ming Zhao
Short	BigD574 "Managing Hot Metadata for Scientific Workflows on Multisite Clouds" Luis Pineda-Morales, Ji Liu, Alexandru Costan, Esther Pacitti, Gabriel Antoniu, Patrick Valduriez, and Marta Mattoso
Short	BigD216 "A Popularity-aware Cost-effective Replication Scheme for High Data Durability in Cloud Storage" Jinwei Liu and Haiying Shen
Short	BigD309 "SLA-Based Profit Optimization for Resource Management of Big Data Analytics-as-a-Service Platforms in Cloud Computing Environments" Yali Zhao, Rodrigo N. Calheros, James, Bailey, Richard Sinnott

S 10: Big Data Applications III	
Short	BigD382 "Exploring Memory Hierarchy and Network Topology for Runtime AMR Data Sharing Across Scientific Applications" Wenzhao Zhang, Houjun Tang, Stephen Ranshous, Surendra Byna, Daniel Martin, Kesheng Wu, Bin Dong, Scott Klasky, and Nagiza Samatova
Short	BigD412 "Estimating Human Interactions with Electrical Appliances for Activity-based Energy Savings Recommendations" Hông-Ân Cao, Tri Kurniawan Wijaya, Karl Aberer, and Nuno Nunes
Short	BigD417 "Application-Driven Sensing Data Reconstruction and Selection Based on Correlation Mining and Dynamic Feedback" Zhichuan Huang, Ting Zhu, and Jianwu Wang
Short	BigD527 "Identifying Dynamic Changes with Noisy Labels in Spatial-temporal Data: A Study on Large-scale Water Monitoring Application" Xiaowei Jia, Xi Chen, Anuj Karpatne, and Vipin Kumar
Short	BigD528 "Optimizing Callout in Unifed Ad Markets" Aman Gupta, S. Muthukrishnan, and Smita Wadhwa
Short	BigD530 "How Interesting Images Are: An Atypicality Approach For Social Networks" Elyas Sabeti and Anders Host-Madsen
Short	BigD599 "Scalable Nearest Neighbor Based Hierarchical Change Detection Framework for Crop Monitoring" Zexi Chen, Ranga Vatsavai, Bharathkumar Ramachandra, Qiang Zhang, Nagendra Singh, and Sreenivas Sukumar
Short	BigD611 "A Scalable Approach for Location-Specific Detection of Santa Ana Conditions" Mai Nguyen, Dylan Uys, Daniel Crawl, Charles Cowart, and Ilkay Altintas
Short	BigD504 "Experiences with Smart City Traffic Pilot" Susanna Pirttikangas, Ekaterina Gilman, Xiang Su, Teemu Leppänen, Anja Keskinarkaus, Mika Rautiainen, Mikko Pyykkönen, and Jukka Riekki

S 11: Big Data Search and Mining in Social Media and Web	
Short	BigD344 "Semi-Supervised Dirichlet-Hawkes Process with Applications of Topic Detection and Tracking in Twitter" Wanying Ding, Yue Zhang, Chaomei Chen, and Xiaohua Hu
Short	BigD395 "Point of Interest Recommendation with Social and Geographical Influence"

	Da-Chuan Zhang, Mei Li, and Chang-Dong Wang
Short	BigD474 "Connection Discovery using Shared Images by Gaussian Relational Topic Model" Xiaopeng Li, Ming Cheung, and James She
Short	BigD633 "What Makes A Group Fail: Modeling Social Group Behavior in Event-Based Social Networks" Xiang Liu and Torsten Suel
Short	BigD637 "Compartmentalized Adaptive Topic Mining on Social Media Streams" Gopi Chand Nutakki and Olfa Nasraoui
Short	BigD410 "Efficient Index Updates for Mixed Update and Query Loads" Sergey Nepomnyachiy and Torsten Suel
Short	BigD444 "Exploiting Temporal Divergence of Topic Distributions for Event Detection" Rongda Zhu, Aston Zhang, Jian Peng, and Chengxiang Zhai
Short	BigD560 "PSH: A Probabilistic Signature Hash Method with Hash Neighborhood Candidate Generation for Fast Edit-Distance String Comparison on Big Data" Joseph Jupin and Yuan Shi
Short	BigD374 "Scalability Analysis of Distributed Search in Large Peer-to-peer Networks" Weimao Ke and Javed Mostafa
Short	BigD533 "Fast Nearest Neighbor Search through Sparse Random Projections and Voting" Ville Hyvönen, Teemu Pitkänen, Sotiris Tasoulis, Elias Jääsaari, Risto Tuomainen, Liang Wang, Jukka Corander, and Teemu Roos

S 12: Data Management & Integration	
Short	BigD559 "RADII: Bridging the Divide between Data and Infrastructure Management to Support Data-Driven Collaborations" Fan Jiang, Claris Castillo, and Charles Schmitt
Short	BigD641 "Advantage of Integration in Big Data: Feature Generation in Multi-Relational Databases for Imbalanced Learning" Farrukh Ahmed, Michele Samorani, Colin Bellinger, and Osmar R. Zaiane
Short	BigD349 "BDTune: Hierarchical Correlation-based Performance Analysis and Rule-based Diagnosis for Big Data System" Rui Ren, Zhen Jia, Lei Wang, Tianxu Yi, and Jianfeng Zhan
Short	BigD324 "Cleaning Antipatterns in an SQL Query Log" Natalia Arzamasova, Martin Schäler, and Klemens Böhm
Short	BigD629 "TruthCore: Non-parametric Estimation of Truth from a Collection of Authoritative Sources" Tathagata Mukherjee, Biswas Parajuli, Piyush Kumar, and Eduardo Pasiliao
Short	BigD422 "Online Multi-view Clustering with Incomplete Views" Weixiang Shao, Lifang He, Chun-Ta Lu, and Philip S. Yu
Short	BigD332 "VHT: Vertical Hoeffding Tree" Nicolas Kourtellis, Gianmarco De Francisci Morales, Albert Bifet, and Arinto Murdopo

Industry and Government Paper Presentations

I&G-regular1: Big Data Analytics	
Regular	N219-Storytelling in Heterogeneous Twitter Entity Network based on Hierarchical Cluster Routing Xuchao Zhang, Zhiqian Chen, Weisheng Zhong, Arnold P. Boedihardjo, and Chang-Tien Lu
Regular	N222-An Architecture for the Deployment of Statistical Models for Big Data Era Juergen Heit, Jiayi Liu, and Mohak Shah
Regular	N226-A Diversified Trending Topic Discovery System Hui Wu, Yi Fang, Huming Wu, and Shenhong Zhu
Regular	N228-Pitfalls of Long-Term Online Controlled Experiments Pavel Dmitriev, Brian Frasca, Somit Gupta, Ron Kohavi, and Garnet Vaz
Regular	N240- LogProv: Logging Events as Provenance of Big Data Analytics Pipelines with Trustworthiness Guoqiang Li, Ruoyu Wang, Daniel Sun, Muhammad Atif, and Surya Nepal

I&G-regular2: Big Data Applications (1)	
Regular	N246-Quantifying Skill Relevance to Job Titles Wenjun Zhou, Yun Zhu, Faizan Javed, Mahmudur Rahman, Janani Balaji, and Matt McNair,
Regular	N227-Fast, Lenient and Accurate: Building Personalized Instant Search Experience at LinkedIn Ganesh Venkataraman, Abhimanyu Lad, Lin Guo, and Shakti Sinha
Regular	N210- Empirical Evaluations of Preprocessing Parameters' Impact on Predictive Coding's Effectiveness Nathaniel Huber-Fliflet, Jianping Zhang, Haozhen Zhao, Robert Keeling, and Rishi Chhatwal
Regular	N216 - Information Retrieval, Fusion, Completion, and Clustering for Employee Expertise Estimation Raya Horesh, Kush Varshney, and Jinfeng Yi
Regular	N225 - Dynamic Pattern Recognition and Classification of HVAC Faults in Commercial Buildings Bradford Littooy, Sophie Loire, Michael Georgescu, and Igor Mezic

I&G-short1: Big Data Algorithms & Systems	
Short	N218-Managing a Complicated Workflow based on Dataflow-based Workflow Scheduler Teruyoshi Zenmyo, Satoshi Iijima, and Ichiro Fukuda
Short	N220- Company Recommendation for New Graduates via Implicit Feedback Multiple Matrix Factorization with Bayesian Optimization Issei Sato, Masahiro Kazama, Haruaki Yatabe, Tairiku Ogihara, Tetsuro Onishi, and Hiroshi Nakagawa
Short	N231- Hermes: A distributed-messaging tool for NLP Ilaria Bordino, Andrea Ferretti, Marco Firrincieli, Francesco Gullo, Marcello Paris, Stefano Pascolutti, and Gianluca Sabena
Short	N232-Cross-Modal Event Summarization: A Network of Networks Approach Jiejun Xu, Samuel Johnson, and Kang-Yu Ni
Short	N237-UStore: An Optimized Storage System for Enterprise Data Warehouses at UnionPay Hongfeng Chai, Hao LIU, Xibo Zhou, Yanjun Xu, Shuo He, Jinzhi Hua, Dongjie He, and Weihuai Liu
Short	N249-An Edge-Set Based Large Scale Graph Processing System Li Zhou, Yinglong Xia, Hui Zang, Jian Xu, and Mingzhen Xia
Short	N256- Automated Port Traffic Statistics: From Raw Data to Visualisation Luca Cazzanti, Antonio Davoli, and Leonardo Millefiori
Short	N259- Knowledge Discovery in Data Science: KDD Meets Big Data Nancy Grady
Short	N262- Forecasting Squatting of Demand in Display Advertising Amita Gajewar, Jignesh Parmar, Lizhong Wu, and Ramana Yerneni,
Short	N264- Human Network Usage Patterns Revealed by Telecom Data Yiming Kong, Hui Zang, and Xiaoli Ma
Short	N243- Classification of Massive Mobile Web Log URLs for Customer Profiling & Analytics

	Rajaraman Kanagasabai, Anitha Veeramani, Hu Shangfeng, Kajanan Sangaralingam, Ying Li, and Giuseppe Manai
Short	N251- A distributed approach to estimating sea port operational regions from lots of AIS data Leonardo Millefiori, Dimitrios Zissis, Luca Cazzanti, and Gianfranco Arcieri

I&G-regular3: Big Data Platforms & Frameworks	
Regular	N208 - Data-at-Rest Security for Spark Syed Yousaf Shah, Brent Paulovicks, and Petros Zerfos
Regular	N214 - SmartCache: Application Layer Caching to Improve Performance of Large-scale Memory Mapping Zhenyun zhuang, Haricharan Ramachandra, Badri Sridharan, Brandon Duncan, Kishore Gopalakrishna, and Jean-Francois Im
Regular	N224- Deep Parallelization of Parallel FP-Growth Using Parent-Child MapReduce Adetokunbo Makanju, Zahra Farzanyar, Aijun An, Nick Cercone, Zane Hu, and Yonggang Hu
Regular	N257- Mini-Apps for High Performance Scientific Data Analysis Sreenivas Sukumar, Michael Matheson, Ramakrishnan Kannan, and Seung-Hwan Lim,
Regular	N260-The state of SQL-on-Hadoop in the Cloud Nicolas Poggi, Josep Ll. Berral, David Carrera, Jose Blakeley, and Nikola Vujic

I&G Panel Session
Big Data Regional Innovation Hubs: Accelerating the Big Data Innovation Ecosystem
Moderator:
Dr. Lea Shanley, co-Executive Director, South Big Data Hub @ RENCI, University of North Carolina-Chapel Hill

I&G-short2: Massive Processing & Experience	
Short	N211-Data Quality: Experiences and Lessons from Operationalizing Big Data Archana Ganapathi and Yanpei Chen
Short	N221-QED: Groupon's ETL management and curated feature catalog system for machine learning Derrick C. Spell, Ling-Yong Wang, Richard Shomer, Bahador Nooraei, Jarrell Waggoner, Xiao-Han Zeng, Jae Chung, Kai-Chen Cheng, and Daniel Kirsche
Short	N223-Uniformization, organization, association and use of metadata from multiple content providers and manufacturers: A close look at the Building Automation System (BAS) sector. Thibaud Nesztler, Don Kasper, Michael Georgescu, Sophie Loire, and Igor Mezic
Short	N245-Extensive Large-Scale Study of Error Surfaces in Sampling-Based Distinct Value Estimators for Databases Vinay Deolalikar
Short	N252- Big-data- driven Anomaly Detection in Industry (4.0): an approach and case study Ljiljana Stojanovic, Marko Dinic, Nenad Stojanovic, and Aleksandar Stojadinovic

I&G-regular4: Big Data Applications (2)	
Regular	N235- Automatic Generation of Relational Attributes: An Application to Product Returns Michele Samorani, Farrukh Ahmed, and Osmar Zaiane
Regular	N239-Predicting Annual Average Daily Highway Traffic from Large Data and Very Few Measurements Tomasz Tajmajer, Malwina Spławińska, Piotr Wasilewski, and Stan Matwin
Regular	N250-Do We Trust Image Measurements? Mylene Simon, Joe Chalfoun, Mary Brady, and Peter Bajcsy
Regular	N258- Detecting Fraud, Corruption, and Collusion in International Development Contracts Elissa Redmiles, Emily Grace, Ankit Rai, and Rayid Ghani,
Regular	N267-Hidden Markov Based Anomaly Detection for Water Supply Systems Zahra Zohrevand, Uwe Gl ässer, Hamed Yaghoubi Shahir, and Mohammad A. Tayebi,

Panels

Panel 1: Big Data and Privacy

We are entering the big data era. We continue to witness the growth of both raw data and data generated from social media and machine learning models. With the emergence of internet of smart things, the data will continue to be the No. 1 runner in the cyberspace, out-grow hardware, software and information technology.

In this panel, the panelists will debate on three most frequently asked questions related to big data and privacy:

(1) Is big data the biggest threat to privacy?
(2) What are the most promising technological solutions for protecting privacy?
(3) Is privacy really dead in the Big Data era?
(4) Is privacy preserving machine learning realistic?

There are several angles for the panelists to share their view points. For example, there is a growing trend for "creating a market" for personal information in which there are explicitly defined rewards for people to share their information (the rewards can be $$$ or in kind). The panelists will engage discussions and debates by considering big data sources, and regulators, such as government, social media, advertisers, data brokers, big data and analytics, ourselves.

Panelists:

Moderator: Ling Liu, Georgia Institute of Technology

1) Kathy Grise, IEEE Big Data Initiative Director
2) George Karypis, University of Minnesota
3) James Joshi, Univ. Pittsburg
4) Ravi Sandhu, Univ. Texas, San Antonio

Bios of Moderator and Panelists

Moderator:

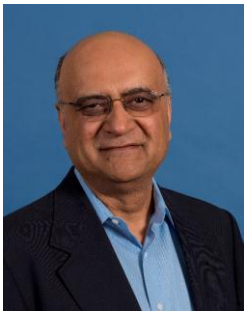
Ling Liu is a Professor in the School of Computer Science at Georgia Institute of Technology. She directs the research programs in Distributed Data Intensive Systems Lab (DiSL), examining various aspects of large scale data intensive systems, including performance, availability, security and privacy. Prof. Liu is an elected IEEE Fellow, a recipient of IEEE Computer Society Technical Achievement Award in 2012. She has published over 300 international journal and conference articles and is a recipient of the best paper award from a dozen of top venues, including ICDCS 2003, WWW 2004, 2005 Pat Goldberg Memorial Best Paper Award, IEEE Cloud 2012, IEEE ICWS 2013, ACM/IEEE CCGrid 2015. In addition to service as general chair and PC chairs of numerous IEEE and ACM conferences in data engineering, very large databases, distributed computing, cloud computing, big data fields, Prof. Liu has served on editorial board of over a dozen international journals. Currently Prof. Liu is the editor in chief of IEEE Transactions on Service Computing. Prof. Liu's current research is primarily sponsored by NSF, IBM and Intel.

Panelists:

Kathy Grise, Senior Program Director - IEEE Future Directions, supports new technology initiatives, and is the IEEE staff program director for the Big Data Initiative, Smart Materials Initiative, the IEEE Technology Navigator, and manages the digital presence team for Future Directions. Prior to joining the IEEE staff, Ms. Grise held numerous positions at IBM, and most recently was a Senior Engineering Manager for Process Design Kit Enablement in the IBM Semiconductor Research and Development Center. Ms. Grise led the overall IT infrastructure implementation, and software development in support of semiconductor device modeling verification, packaging, and delivery; device measurement and characterization data collection and management, and automation for device modeling engineers. Ms. Grise is a graduate of Washington and Jefferson College, and an IEEE Senior member.

James Joshi is a professor of School of Information Sciences (SIS) at the University of Pittsburgh. He received his MS in Computer Science and PhD in Computer Engineering degrees from Purdue University in 1998 and 2003, respectively. He is an elected Fellow of the Society of Information Reuse and Integration (SIRI) and is an Senior member of IEEE and ACM. His research interests include Access Control Models, Security and Privacy of Distributed Multimedia Systems, Trust Management and Information Survivability. He is a recipient of the NSF-CAREER award in 2006. His current research activities include Security and Privacy issues in Cloud Computing and Collaborative systems, Social networks (including access control anonymization techniques and privacy attack analysis), and Intrusion detection/prevention systems in networks (Trust based approaches, collaborative detection, etc.). At Pitt, he co-founded and is the director of the Laboratory of Education and Research in Security Assured Information Systems (LERSAIS), which is one of only about dozen in the nation with five CNSS certifications, and manages the DoD Information Assurance Scholarship Program and the NSF-Federal Cyber Service Scholarship for Service program. His efforts has resulted in the CAE and CAE-Research designation of LERSAIS jointly by NSA and the DHS.

George Karypis is a Distinguished McKnight University Professor and an ADC Chair of Digital Technology at the Department of Computer Science & Engineering at the University of Minnesota, Twin Cities. His research interests spans the areas of data mining, high performance computing, information retrieval, collaborative filtering, bioinformatics, cheminformatics, and scientific computing. His research has resulted in the development of software libraries for serial and parallel graph partitioning (METIS and ParMETIS), hypergraph partitioning (hMETIS), for parallel Cholesky factorization (PSPASES), for collaborative filtering-based recommendation algorithms (SUGGEST), clustering high dimensional datasets (CLUTO), finding frequent patterns in diverse datasets (PAFI), and for protein secondary structure prediction (YASSPP). He has coauthored over 260 papers on these topics and two books ("Introduction to Protein Structure Prediction: Methods and Algorithms" (Wiley, 2010) and "Introduction to Parallel Computing" (Publ. Addison Wesley, 2003, 2nd edition)). In addition, he is serving on the program committees of many conferences and workshops on these topics, and on the editorial boards of the IEEE Transactions on Knowledge and Data Engineering, ACM Transactions on Knowledge Discovery from Data, Data Mining and Knowledge Discovery, Social Network Analysis and Data Mining Journal, International Journal of Data Mining and Bioinformatics, the journal on Current Proteomics, Advances in Bioinformatics, and Biomedicine and Biotechnology.



Ravi Sandhu is founding Executive Director of the Institute for Cyber Security at the University of Texas San Antonio, and holds an Endowed Chair. He is an ACM, IEEE and AAAS Fellow and inventor on 30 patents. He has received the IEEE Computer Society Technical Achievement award, and the ACM SIGSAC outstanding innovation and outstanding contribution awards. He is past Editor-in-Chief of the IEEE Transactions on Dependable and Secure Computing, past founding Editor-in-Chief of ACM Transactions on Information and System Security and a past Chair ofACM SIGSAC. He founded ACM CCS, SACMAT and CODASPY, and has been a leader in numerous other security research conferences. His research has focused on security models and architectures, including the seminal role-based and attribute-based access control models, and their applications in cloud, mobile and social computing. His papers have over 32,000 Google Scholar citations including over 7,600 for the RBAC96 paper.

Panel 2: Big Data in Industry and Government - Challenges, Requirements and Constraints

This panel will focus on understanding the differences in Big Data related challenges, requirements and constraints in the industry from those in the government sector. Panelists will share their Big Data use in their organizations. They will discuss the requirements and constraints that drive Bid Data analytics and technologies in these two sectors. They will explore commonalities and differences so as to understand how industry and government sectors can seek to support each other's needs and work more closely, including sharing of resources and

Panelists:

Moderator: Eui-Hong (Sam) Han, Director, Big Data & Personalization, The Washington Post, Washington, DC

1) Kristen Summers, IBM
2) Andrew Schechtman-Rook, Capital One
3) Raju Vatsavai, NC State University
4) Aidong Zhang, SUNY Buffalo & NSF
5) Jan Neumann, Comcast

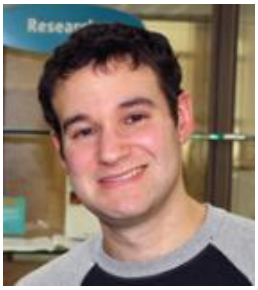
Bios of Moderator and Panelists

Moderator:

Eui-Hong (Sam) Han is the Director, Big Data & Personalization at The Washington Post. Sam is leading a team to build an integrated Big Data platform to store all aspects of customer profiles and activities from both digital and print circulation, metadata of content, and business data. Utilizing the integrated data from the platform, his team builds tools and services to provide personalized experience to customers, to empower newsroom with data for better decisions, and to provide targeted advertising capability. His expertise includes data mining, machine learning, information retrieval, and high performance computing. He holds PhD in Computer Science from the University of Minnesota.

Panelists:

Kristen Summers is the Technical Delivery Lead for Watson Implementations in the Public Sector at IBM. She has been working in technologies for processing unstructured textual data, with a primary focus on applied research, for the past 15 years. Her experience in this area includes leading projects on, entity extraction, entity co-reference, machine translation, and other related topics. Before joining IBM Watson, she was Technical Director in the Knowledge and Information Management Division Group at CACI. She holds a PhD in Computer Science from Cornell University and a BA in Computer Science and English from Amherst College.



Andrew Schechtman-Rook is a data scientist at Capital One, primarily trying to figure out how to build and deploy models in reusable, robust, and data scientist-friendly ways. Before going corporate Andrew got his PhD in Astronomy at the University of Wisconsin-Madison, blending cutting-edge modeling tools and detailed observations to better understand how stars are distributed in galaxies. In his increasingly limited spare time he analyzes NFL statistics, writing up the most interesting results on his blog at http://phdfootball.blogspot.com.

Raju is a Chancellor's Faculty Excellence Program Geospatial Analytics Cluster Associate Professor in the Department of Computer Science, North Carolina State University (NCSU). He works at the intersection of spatial and temporal big data management, analytics, and high performance computing with applications in the national security, geospatial intelligence, natural resources, climate change, location-based services, and human terrain mapping. Before joining NCSU, Raju was the Lead Data Scientist for the Computational Sciences and Engineering Division (CSED) at the Oak Ridge National Laboratory (ORNL). He has published more than 80 peer-reviewed articles in conferences and journals, and edited two books on "Knowledge Discovery from Sensor Data." He served on program committees of leading international conference including ACM KDD, ACM GIS, ECML/PKDD, SDM, CIKM, IEEE BigData, and co-chaired several workshops including ICDM/SSTDM, ICDM/KDCloud, ACM SIGSPATIAL BigSpatial, Supercomputing/BDAC, KDD/LDMTA, KDD/Sensor-KDD, and SDM/ACS. He holds MS and PhD degrees in computer science from the University of Minnesota

Aidong Zhang is currently on leave from the State University of New York (SUNY) at Buffalo and serving as a program director in the Information & Intelligent Systems division of the Directorate for Computer & Information Science & Engineering, National Science Foundation. Dr. Zhang is a SUNY Distinguished Professor of Computer Science and Engineering. Her research interests include data mining/data science, bioinformatics, health Informatics, multimedia and database systems, and content-based image retrieval. She has authored over 290 research publications in these areas. She has chaired or served on over 160 program committees of international conferences and workshops, and currently serves on several journal editorial boards. Dr. Zhang is an IEEE Fellow

Jan Neumann is a senior manager at Comcast Labs Washington DC where he leads the research team. His team combines large-scale machine learning, deep learning, NLP and computer vision to develop novel algorithms and product concepts that improve the experience of Comcast's customers. Before Comcast, he worked for Siemens Corporate Research on various computer vision related projects such as driver assistance systems and video surveillance. He has published over 20 paper in scientific conferences and journals, and is a frequent speaker on machine learning and data science. He holds a Ph.D. in Computer Science from the University of Maryland, College Park.

Panel 3: Big Data Regional Innovation Hubs: Accelerating the Big Data Innovation Ecosystem

In recognition of the substantial and growing impact of big data to the U.S., across sectors, in 2012 the White House launched a multi-agency research initiative to foster and coordinate big data innovation across the US. Under this initiative, the National Foundation launched four Big Data Regional Innovation Hubs, new organizations intended to develop the Big Data innovation ecosystem and facilitate thematic communities' use of data sciences for societal benefit. Specifically, the Big Data Regional Innovation Hubs accelerate partnerships among people in business, academia, and government who apply data science and analytics to help solve regional and national challenges. The Big Data (BD) Hubs cover all 50 states and currently include several hundred universities, corporations, federal agencies, and non-governmental organizations. We will introduce the Big Data Hubs and report on some of the significant activities underway in Digital Agriculture, Transportation, Data Infrastructure, Smart Cities, Data Sharing, Privacy and Security. Finally, we will discuss opportunities to engage with the BD Hubs and our growing networks of Public/Private partnerships.

Panelists:

Moderator: Dr. Lea Shanley, co-Executive Director, South Big Data Hub @ RENCI, University of North Carolina-Chapel Hill

1) Dr. Rene Baston, Executive Director, North East Big Data Innovation Hub
2) Dr. Melissa Craigin, Executive Director, Midwest Big Data Innovation Hub
3) Dr. Meredith Lee, Executive Director, West Big Data Innovation Hub
4) Dr. Renata Rawlings-Goss, co-Executive Director, South Big Data Innovation Hub @ GA Tech

Bios of Moderator and Panelists

Moderator:

Dr. Lea Shanley is a founding co-Executive Director of the South Big Data Innovation Hub at the Renaissance Computing Institute (RENCI) at the University of North Carolina-Chapel Hill. Before joining the Hub, Dr. Shanley served as a White House Presidential Innovation Fellow at NASA Headquarters, where she designed and guided open innovation and open source research strategies for planetary and Earth science. From 2013 to 2015, Dr. Shanley founded and led the Federal Crowdsourcing and Citizen Science Community of Practice, growing the community to more than 300 members from 40 agencies, and advising and leading the development of the online citizen science toolkit, which became CitizenScience.gov. From 2011-2014, Dr. Shanley was the founding director of the Washington-based Wilson Center Commons Lab, guiding strategic research in crowd-mapping, social computing, and big data, and conceptualizing and initiating the federal citizen science toolkit and projects database. Previously, she served as an American Association for the Advancement of Science/ASA-CSSA-SSSA Congressional Science Fellow and primary science advisor to Senator Bill Nelson (FL), where she made significant contributions to the NASA Authorization Act of 2010 and two other statutes. Dr. Shanley also helped to launch the Wisconsin Geographic Information Coordination Council, and spent more than 15 years conducting research and working with local, state, and tribal governments in the development of GIS-based decision support systems for city planning, environmental monitoring, coastal management, and disaster response. She holds a Ph.D. in Environment and Resources, with a focus on geographic information science and remote sensing, at the University of Wisconsin-Madison. Her work has been featured by whitehouse.gov, Fast Company, Popular Science, The Washington Post, NextGov, TechCrunch, and Vice.

Panelists:

Dr. René Baston is the Executive Director of the Northeast Big Data Innovation Hub. He has over 20 years of experience in innovation, business development, public/private collaborations, entrepreneurship, technology transfer & commercialization, consulting, and economic development. He is currently the Executive Director of the Northeast Big Data Innovation Hub, launched by the National Science Foundation, focused on identifying substantial societal challenges and building multi-sector, multi-disciplinary partnerships to address them with data-driven solutions. He is also an instructor of the Lean Startup methodology for the NYC Regional Innovation Node and the co-founder of three startups. Previously, he was the Special Advisor on Innovation and Entrepreneurship to the City University of New York Vice Chancellor for Research; the Director of Industry Interactions and Entrepreneurship at the Columbia University Data Science Institute; the Chief Business Officer at the New York Academy of Sciences, for which he led all aspects of global business development, business strategy, and strategic initiatives; Associate Director at Columbia's Science & Technology Ventures, one of the world's leading academic technology transfer organizations; a management consultant in Ernst & Young's healthcare/HIT consulting group; and spent several years in the laboratory of Nobel Laureate, Eric Kandel, at the Columbia University Center for Neurobiology and Behavior. He earned both his Master's in Biomedical Informatics and his B.A. from Columbia University.

Dr. Melissa Cragin is the Executive Director for the Midwest Big Data Hub, based at the University of Illinois at Urbana-Champaign (UIUC) in the National Center for Supercomputing Applications (NCSA). Prior to joining NCSA, Melissa was Staff Associate in the Office of the Assistant Director, Directorate of Biological Sciences at the National Science Foundation (NSF), where she guided the development of data policy and accelerated community engagement on research data management and public access. Before joining the staff at NSF, Melissa served for two years in the BIO Directorate as an AAAS Science & Technology Policy Fellow. Previous to her work with the federal government, Melissa was on the faculty of the Graduate School of Library and Information Science at the University of Illinois, where she led the Data Curation Education Program and conducted research in the Center for Informatics Research in Science and Scholarship. She has a PhD from UIUC and an MLIS from Rutgers University.

Dr. Meredith Lee is the Executive Director of the West Big Data Innovation Hub, a consortium launched by the National Science Foundation to address societal challenges with Big Data innovation. The West Hub is led by UC Berkeley, UC San Diego, and the University of Washington, and includes Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming. Dr. Lee previously served as a Science & Technology Policy Fellow at the U.S. Department of Homeland Security (DHS) Homeland Security Advanced Research Projects Agency (HSARPA), guiding strategic research in graph analytics, risk assessment, machine learning, data visualization, and distributed computing. She co-led the White House Innovation for Disaster Response and Recovery Initiative as well as the Ideation Community of Practice, a network of Federal innovators from more than 25 agencies. Meredith completed her Ph.D. in Electrical Engineering at Stanford University and was a postdoctoral researcher at the Canary Center for Cancer Early Detection. She was previously at MIT Lincoln Laboratory, Intel, IBM T.J. Watson Research Center, and Agilent Laboratories. Dr. Lee is a co-founder of NationOfMakers.org, past president of the Stanford Optical Society of America/SPIE, and served on the first Steering Committee for the National Photonics Initiative. Her work has been featured by whitehouse.gov, Make, ArsTechnica, The Washington Post, Forbes, and Fast Company.

Dr. Reneata Rawlings-Goss, co-Executive Director, South Big Data Innovation Hub @ GA Tech

Special Sessions

SPECIAL SESSION I: 2nd SPECIAL SESSION ON INTELLIGENT DATA MINING

Session Organizer: Uraz Yavanoglu, PhD

Summary:
Recent developments in processing, storing, and sharing huge amount of data become problematic due to the lack of new approaches, techniques, methods, algorithms and technologies. Researchers try to find proper solutions based on their experiences and make contributions to current data mining and classification knowledge. This approach actually causes new problems due to the missing theoretical notions, lack of necessary disciplines and insufficient awareness on data security, information retrieval, social networking within behavioral and social issues on human nature. This special session traces the gap between big data, artificial intelligence (AI), machine learning (ML) and data mining.

Nowadays, researchers use interdisciplinary way to understand knowledge among all types of resource including data, document, tool, device, experience, process and people. This approach may help to understand biological evolution to propose robust and powerful approaches between human nature and big data processing.

Intelligent Data Mining term is not only related to Computer Science. This special session opens to every researcher as well as industrial partners to make contributions.

Short Bio.
Uraz Yavanoglu was born in Ankara, capital city of Turkey. He is a researcher at Gazi University Department of Computer Engineering. He received his M.Sc. degree from Gazi University Department of Computer Engineering and Ph.D. from Gazi University Faculty of Technology. His research interests are Artificial Intelligence, Data Mining, Information Security, Forensic Analysis and Computer Graphics. He was leading researcher of Gazi University Technology and Innovation Center (GUTIC). He received TUBITAK post doc scholar in 2014. He had completed his post-doctoral research at Arizona State University, School of Computing, Informatics, and Decision Systems Engineering. He is member of ASU Data Mining and Machine Learning Laboratory and AZComp Fellows.

Schedule **December 6 Tuesday 2016**		
Time	**Title**	**Presenter/Author**
07:30am-8:00am	**Registration**	
08:30am-08:45am	**Session Opening Speech**	
	Uraz Yavanoglu, PhD	
08:45am-09:00am	Subgroup Discovery on Big Data: pruning the search space on exhaustive search algorithms	Francisco Padillo, José María Luna, and Sebastián Ventura
09:00am-09:15am	MapReduce-based Deep Learning With Handwritten Digit Recognition Case Study	Nada Basit, Yutong Zhang, Hao Wu, Haoran Liu, Jieming Bin, Yijun He, and Abdeltawab Hendawi
09:15am-09:30am	From Big Data to Big Challenge: An optimized frequent pattern mining algorithm with multiple minimum supports	Hsiao-Wei Hu, Hao-Chen Chang, and Wen-Shiu Lin
09:30am-09:45am	Swarm Intelligence (SI) based Profiling and Scheduling of Big Data Applications	Thamarai Selvi Somasundaram, Kannan Govindarajan, and Vivekanandan Suresh Kumar
09:45am-10:00am	Online Trajectory Segmentation and Summary With Applications to Visualization and Retrieval	Yehezkel Resheff
10:00am-10:15am	Skeleton Decomposition Analysis for Subspace Clustering	Ali Sekmen, Akram Aldroubi, Ahmet Bugra Koku, and Ahmet Faruk Cakmak

10:15am-10:30am	Big Data Analytics in Cloud Gaming: Players' Patterns Recognition using Artificial Neural Networks	Victor Perazzolo Barros and Pollyana Notargiacomo
10:30am-10:45am	Identifying Trolls and Determining Terror Awareness Level in Social Networks Using a Scalable Framework	Busra Mutlu, Merve Mutlu, Kasim Oztoprak, and Erdogan Dogdu
10:45am-11:00am	**Break**	
11:00am-11:15am	Classifying Tables in Financial Documents	Quanzhi Li
11:15am-11:30am	Urban Human Mobility Data Mining: An Overview	Kai Zhao, Sasu Tarkoma, Siyuan Liu, and Huy Vo
11:30am-11:45am	A Real-Time Autonomous Highway Accident Detection Model Based on Big Data Processing and Computational Intelligence	Murat Ozbayoglu, Gokhan Kucukayan, and Erdogan Dogdu
11:45am-12:00pm	An Extended IoT Framework with Semantics, Big Data, and Analytics	Omer Berat Sezer, Erdogan Dogdu, Murat Ozbayoglu, and Aras Can Onal
12:00pm-12:15pm	An Overview Of Studies About Students' Performance Analysis and Learning Analytics in MOOCs	ismail duru, Gülüstan Doğan, and Banu Diri
12:15pm-12:30pm	A Survey on Semantic Web and Big Data Technologies for Social Network Analysis	Sercan Kulcu and Erdogan Dogdu
12:30pm-12:45pm	Patient-Record Level Integration of De-Identified Healthcare Big Databases	Xiao Li, Reza Sharifi Sedeh, Liao Wang, and Yang Yang
12:45pm-02:00pm	**Lunch Break**	
02:00pm-02:15pm	Improving Clustering Efficiency by SimHash-based K-Means Algorithm for Big Data Analytics	Jenq-Haur Wang and Jia-Zhi Lin
02:15pm-02:30pm	The Difference-of-Datasets Framework: A Statistical Method to Discover Insight	Paul Raff and Ze Jin
02:30pm-02:45pm	Improving item-based recommendation accuracy with user's preferences on Apache Mahout	Ammar Jakabji and Hasan Dağ
02:45pm-03:00pm	Solving Cold Start Problem in Large-scale Recommendation Engines: A Deep Learning Approach	Jianbo Yuan, Walid Shalaby, Mohammed Korayem, David Lin, and Khalifeh AlJadda
03:00pm-03:15pm	User and Entity Behavior Analytics for Enterprise Security	Madhu Shashanka, Min-Yi Shen, and Jisheng Wang
03:15pm-03:30pm	Event Detection from Social Network Streams Using Frequent Pattern Mining with Dynamic Support Values	Nora Alkhamees and Maria Fasli
03:30pm-03:45pm	Smart Online Vehicle Tracking System for Security Applications	Brahim Hnich, Ata Sasmaz, Özkan Sayın, Faisal R. Al-Osaimid, and Amine Lamine
03:45pm-04:00pm	Change Detection and Classification of Digital Collections	Sampath Jayarathna and Faryaneh Poursardar
04:00pm-04:30pm	**Break**	
04:30pm-04:45pm	DelayRadar: A Multivariate Predictive Model for Transit Systems	Aparna Oruganti, Fangzhou Sun, Hiba Baroud, and Abhishek Dubey
04:45pm-05:00pm	Text Mining and Sentiment Extraction in Central Bank Documents	Giuseppe Bruno
05:00pm-05:15pm	The Effect of Pets on Happiness: A Data-Driven Approach via Large-Scale Social Media	Yuchen Wu, Jianbo Yuan, Quanzeng You, and Jiebo Luo

05:15pm-05:30pm	Fine-grained Mining of Illicit Drug Use Patterns Using Social Multimedia Data from Instagram	Yiheng Zhou, Numair Sani, and Jiebo Luo
05:30pm-05:45pm	A feature selection method based on Lorentzian metric	Yerzhan Kerimbekov and Hasan Şakir Bilge
05:45pm-06:00pm	Intelligent Authorship Identification with using Turkish Newspapers Metadata	Ozlem Yavanoglu
06:00pm-06:15pm	A Bayesian Predictor of Airline Class Seats Based on Multinomial Event Model	Bingchuan Liu, Yudong Tan, and Huimin Zhou
06:15pm-06:30pm	To Link or Not to Link: Ranking Hyperlinks in Wikipedia using Collective Attention	Jaroslav Cechak, Philip Thruesen, Blandine Seznec, Roel Castano, and Nattiya Kanhabua
06:30pm-06:45pm	Event Segmentation using Parallel MRK-Means Clustering based on MapReduce	Omair Shafiq

Intelligent Data Mining Session Program Committee		
Member	**Affiliation**	**Country**
Abdulkerim Senoglu	Gazi University	Turkey
Adnan Ozsoy	Hacettepe University	Turkey
Arzucan Ozgur	Bogazici University	Turkey
Aysenur Birturk	Middle East Technical University	Turkey
Begum Mutlu	Gazi University	Turkey
Burcu Can	Hacettepe University	Turkey
Ebru Aydogan	Gazi University	Turkey
Engin Demir	Cankaya University	Turkey
Erdogan Dogdu	TOBB University of Economics and Technology	Turkey
Erman Ayday	Bilkent University	Turkey
Feyza Yildirim Okay	Gazi University	Turkey
Gokberk Cinbis	Bilkent University	Turkey
Gulfem Isiklar Alptekin	Galatasaray University	Turkey
Gunce Orman	Galatasaray University	Turkey
H. Altay Guvenir	Bilkent University	Turkey
Hasan Şakir Bilge	Gazi University	Turkey
Hayri Sever	Hacettepe University	Turkey
Ilhami Colak	Nisantasi University	Turkey
Jiebo Luo	Rochester University	USA
Kai Shu	Arizona State University	USA
Kamer Kaya	Sabanci University	Turkey
Kannan Govindarajan	Athabascau University	Canada
Kasim Oztoprak	Karatay University	Turkey
Lillian Ratliff	University of Washington	USA
M. Sedef Demirci	Gazi University	Turkey
Magdalini Eirinaki	San Jose State University	USA
Mahmut Kaya	Gazi University	Turkey
Mehmet Demirci	Gazi University	Turkey

Muhammet Unal	Gazi University	Turkey
Murat Haciomeroglu	Gazi University	Turkey
Murat Ozbayoglu	TOBB University of Economics and Technology	Turkey
Oktay Yildiz	Gazi University	Turkey
Ramazan Bayindir	Gazi University	Turkey
Seref Sagiroglu	Gazi University	Turkey
Taflan Gundem	Bogazici University	Turkey
Tolga Ensari	Istanbul University	Turkey
Tunga Gungor	Bogazici University	Turkey
Yan Wan	University of North Texas	USA

SPECIAL SESSION II: Granular Computing in Big Data

Summary:

It is our pressure to welcome you to this special session. Superficially granular computing (GrC) means "granulate and compute". It is a generalization of classical divide and conquer. To divide, one may directly granulate the Turing machine or indirectly granulate the data. A granule of data is a piece of knowledge or a "seed" of uncertainty. Historically, the concept of granular computing is derived from granular mathematics. The central points in GrC have been concerning the concepts of granule, knowledge, uncertainty, implicitly complexity and parallelism; Big Data shares the same issues. Plainly Big Data can be the playpen for GrC. A granule is often more than a subset. In the language of computer science, a granule is a variable that takes values in subsets of the universe U. Mathematically the domain of such a variable is the granule, which is a member of P(P(U)), the double power set of U; note that by regarding x as {x}, P(U) can be regarded as a subset of P(P(U)). We will adopt both views; both variables and domains are granules. For examples, the topological neighborhood system of a point is a granule. Of course, by our convention, a granule can be a subset of U. A class of very powerful examples is the class of simplicial complexes in algebraic topology.

Finally I would like to express many thanks to conference organizers for their supports. Most importantly, we thank all the authors for their contributions.

Tsau-Young Lin
Nov 20, 2016

Dec 8, 2016 8:00-10:45AM		
1	SP02210	Tsau-Young Lin, *Very Fast Frequent Itemset Mining Simplicial Complex Methods* **Author Email(s):** ty.lin@sjsu.edu **Contact Person:** Tsau-Young Lin <ty.lin@sjsu.edu>
2	SP02204	Hiroshi Sakai, Chenxi Liu, Michinori Nakata, and Shusaku Tsumoto, *A Proposal of a Privacy-preserving Questionnaire by Non-deterministic Information and Its Analysis* **Author Email(s):** sakai@mns.kyutech.ac.jp, p350932s@mail.kyutech.jp, nakatam@ieee.org, tsumoto@med.shimane-u.ac.jp **Contact Person:** Hiroshi Sakai <sakai@mns.kyutech.ac.jp>
3	SP02202	Parul Sharma and Teng-Sheng Moh, *Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter* **Author Email(s):** teng.moh@sjsu.edu **Contact Person:** Teng-Sheng Moh <teng.moh@sjsu.edu>
4	SP02206	Shusaku Tsumoto, Shoji Hirano, and Haruko Iwata, *Construction of Clinical Pathway from Histories of Clinical Actions in Hospital Information System* **Author Email(s):** tsumoto@med.shimane-u.ac.jp, hirano@med.shimane-u.ac.jp **Contact Person:** Shusaku Tsumoto <tsumoto@med.shimane-u.ac.jp>
5	SP02208	Yan Zhu, Melody Moh, and Teng-Sheng Moh, *Multi-Layer Text Classification with Voting for Consumer Reviews* **Author Email(s):** teng.moh@sjsu.edu, melody.moh@sjsu.edu **Contact Person:** Teng-Sheng Moh <teng.moh@sjsu.edu>
6	SP02207	Shusaku Tsumoto, Shoji Hirano, Haruko Iwata, and Tomohiro Kimura, *Mining Process for Improvement of Clinical Process Quality* **Author Email(s):** tsumoto@med.shimane-u.ac.jp, hirano@med.shimane-u.ac.jp, haruko23@med.shimane-u.ac.jp, t-kimura@med.shimane-u.ac.jp **Contact Person:** Shusaku Tsumoto <tsumoto@med.shimane-u.ac.jp>
7	SP02201	Zhenwei Du and Haopeng Chen, *Research on the big data system of massive open online course* **Author Email(s):** huoganlan@sjtu.edu.cn, chen-hp@sjtu.edu.cn **Contact Person:** Zhenwei Du <huoganlan@sjtu.edu.cn>
8	SP02205	Srinivasa Rao Kundeti, Vijayananda J, Srikanth Mujjiga, and Kalyan Chakravarthi Murahari, *Clinical Named Entity Recognition: Challenges and opportunities* **Author Email(s):** srinivasa.rao@philips.com, vijayananda.j@philips.com, srikanth.mujjiga@philips.com, kalyan.chakravarthi.murahari@philips.com **Contact Person:** Srinivasa Rao Kundeti <srinivasa.rao@philips.com>
9	SP02209	Smrithy G S and Ramadoss Balakrishnan, *Online Anomaly Detection using Non-Parametric Technique for Big Data Streams in Cloud Collaborative Environment* **Author Email(s):** smrithygs1990@gmail.com **Contact Person:** Smrithy G S <smrithygs1990@gmail.com>

Manufacturing Symposium

Symposium on Data Analytics for Advanced Manufacturing

Theme: From Sensing to Decision-Making

IEEE Big Data Conference

Dec 5-8, 2016, Washington D.C., USA

Tentative Program:

December 6, 2016

Time	Event
8:45 – 9:45	**Conference Keynote Speech:** *Database Decay and How to Avoid It* **Dr. Michael Stonebraker, Paradigm4/MIT, USA**
9:45 – 10:45	**Conference Keynote Speech:** *Leveraging High Performance Computing to Drive Advanced Manufacturing R&D at the US Department of Energy* **Mark Johnson, Advanced Manufacturing Office, U.S. Department of Energy**
10:45 – 11:05	*Coffee Break*
11:05 – 11:15	**Opening Remarks: Sudarsan Rachuri, DOE**
11:15 – 12:00	**Symposium Keynote Speech:** **Dr. Frank W. Gayle, Advanced Manufacturing National Program Office (AMNPO), NIST**
12:00 – 12:45	**Symposium Keynote Speech:** *The GE Brilliant Factory* **Dr. Matteo Bellucci, GE Global Research Center, Niskayuna, NY.**
12:45– 14:00	*Lunch*
14:00 – 16:05	**Panel: Big Data Analytics for Advanced Manufacturing: Challenges and opportunities** *Panelists:* Dr. Matteo Bellucci (GE), Dr. Ram Sriram (NIST), Matthew Jacobsen (USAF), Prof. Sankaran Mahadevan (Vanderbilt University), Prof. Soundar Kumara (Penn State), Dr. Valerie R. Coffman (Xometry), Dr. Sivaramakumar Gopalasundaram (Cognizant) *Panel Moderator*: Dr. Sudarsan Rachuri (DOE)
16:05 – 16:25	*Coffee Break*
16:25 – 18:05	**Session 1** *(Session Chair: Dr. Ronay Ak)*
16:25 – 16:50	Max Ferguson, Kincho Law, Raunak Bhinge, Yung-Tsun Tina Lee, and Jinkyoo Park, *Evaluation of a PMML-Based GPR Scoring Engine on a Cloud Platform and Microcomputer Board for Smart Manufacturing*
16:50 – 17:15	Shakti Awaghad, *SCEM: Smart & Effective Crowd Management with a Novel Scheme of Big Data Analytics*
17:15 – 17:40	Dazhong Wu, Connor Jennings, Janis Terpenny, and Soundar Kumara, *Cloud-Based Machine Learning for Predictive Analytics: Tool Wear Prediction in Milling*
17:40 – 18:05	Alexander Brodsky, Mohan Krishnamoorthy, William Bernstein, and M. Omar Nachawati, *A System and Architecture for Reusable Abstractions of Manufacturing Processes*

December 7, 2016

Time	Event
8:45 – 9:45	**Conference Keynote Speech:** *Harnessing the Data Revolution: A Perspective from the National Science Foundation* **Dr. Chaitanya Baru, National Science Foundation**
9:45 – 10:45	**Conference Keynote Speech:** *On the Power of Big Data: Mining Structures from Massive, Unstructured Text Data* **Prof. Jiawei Han, University of Illinois at Urbana-Champaign, USA**
10:45 – 11:05	*Coffee Break*
11:05 – 11:10	**Opening Remarks: Tina Lee, NIST**
11:10 – 11:55	**Symposium Keynote Speech:** **Dr. Rumi Ghosh, Robert Bosch LLC**
11:55 – 12:45	**Session 2** *(Session Chair: Dr. Anantha Narayanan)*
11:55 – 12:20	Srinivasan Radhakrishnan, and Sagar Kamarthi, *Convergence and Divergence in Academic and Industrial Interests on IOT based Manufacturing*
12:20 – 12:45	Srinivasan Radhakrishnan, and Sagar Kamarthi, *Complexity-Entropy Feature Plane for Gear Fault Detection*
12:45 – 14:00	*Lunch*
14:00 – 15:45	**Session 3: Bosch Big Data Challenge** *(Session Chair: Dr. Rumi Ghosh)*
14:00 – 14:05	Introduction to Bosch Data Challenge – Dr. Rumi Ghosh
14:05 – 14:30	Bohdan Pavlyshenko, *Machine Learning, Linear and Bayesian Models for Logistic Regression in the Failure Detection Problems*
14:30 – 14:55	Darui Zhang, Bin Xu, and Jasmine Wood, *Predict Failed Product Using Large-scale Data: A Two-stage Approach with Clustering and Supervised Learning*
14:55 – 15:20	Abhinav Maurya, *Bayesian Optimization for Predicting Rare Internal Failures in Manufacturing Processes*
15:20 – 15:45	Ankita Mangal and Nishant Kumar, *Using Big Data to enhance the Bosch Production Line Performance: A Kaggle Challenge*
15:45 – 16:10	**Invited Talk:** *Advancing Additive Manufacturing Through Visual Data Science* **Dr. Chad Steed, ORNL**
16:10 – 16:25	*Coffee Break*
16:25 – 16:40	**Closing Remarks**
16:40	**Closure and Open Discussion**

Keynote Speeches

Keynote 1:

Speaker: Dr. Frank W. Gayle, Advanced Manufacturing National Program Office (AMNPO), NIST

Bio

Dr. Frank Gayle is Deputy Director of the interagency Advanced Manufacturing National Program Office (AMNPO) which is headquartered at the National Institute of Standards and Technology (NIST). Frank received an Sc.D. in Metallurgy from the Massachusetts Institute of Technology, as well as an M.S. in Mechanical Engineering and Materials Science and a B.S.E. in Civil Engineering, both from Duke University.

Prior to coming to NIST, Dr. Gayle spent 11 years in industry in the field of alloy development for aerospace applications. Before joining the AMNPO in December 2012, Frank worked in the NIST Metallurgy Division in positions from research metallurgist to Division Chief. His research covered a wide range of materials, including quasicrystals, lead-free solder, and aerospace materials, including creating materials for NASA's Space Shuttle and identifying strengthening mechanisms in the original Wright Brothers' Flyer engine. As Division Chief, he was responsible for broad support of industry needs for measurements, standards, and data in the application of metals. As Chief, Frank developed major programs in energy, microelectronics, and metals for mechanical applications, focusing on measurement needs for industry.

From 2002 through 2007, Frank headed the NIST-led team of scientific experts investigating the steel forensics involved in the collapse of the World Trade Center towers during the September 11, 2001 attacks. He has twice won the Department of Commerce Gold Medal, the Department's highest award. As Deputy Director, Frank is responsible for the operations of the AMNPO, and leads efforts to carry out the Congressionally mandated development of the Manufacturing USA program.

Keynote 2:

Title: The GE Brilliant Factory

Speaker: Dr. Matteo Bellucci, GE Global Research Center, NY

Abstract

With the increase in global competition for high-quality products and compressed development schedules due to shortened product lifecycles, Brilliant Factory allows early entry into the marketplace. The use of Brilliant Factory tools is essential for simulating individual manufacturing processes and the total manufacturing system. By driving compatibility between the product design and the manufacturing plants, these virtual tools and methods enable the early optimization of cost, quality, and time to help achieve integrated products, process and resource design, and affordability.

Brilliant Factory envisions an approach to enable digital manufacturing that brings total digital integration within and between every part of the value chain starting from the design phase all the way to supply chain and service. The digital thread connects product development and design, manufacturing system and process design, material flow systems, manual and automated fabrication and assembly processes, quality verification, distribution, service and lifecycle management. The Brilliant Factory concept allows for value to be realized in each of these parts of the value chain, where collaboration between designers, manufacturing engineers and operations, is enabled by a "Digital Thread".

Bio:

Dr. Matteo Bellucci joined GE Oil & Gas, Florence, Italy, in 2007 as NPI Program Manager, taking responsibility for the largest NPI undertaken by O&G at that time. He also led the development of the Boil Off product line. His last assignment was to develop the Train Configuration Tool that enabled to create a technical offer for a full train in hours Vs. weeks.

In 2011 Dr. Bellucci moved to GE Global Research, taking the lead of the Processes and System lab. He and his team are leading most of the Brilliant Factories across the companies, spanning virtual validation of new factories as well as processes such as Casting and Additive. His team also understand how to optimize maintenance, and increase automation and throughput of various manufacturing processes.

Prior to joining GE, Dr. Bellucci worked as Test Engineer in the Icing Wind Tunnel at the Italian Aerospace Research Center where he acted as focal point for the Airbus A380 and A400 Military icing test campaigns.

Dr. Bellucci attended the University of Naples and graduated with a PhD, in Aerospace Engineering. As part of his education Matteo worked at the Von Karman institute of Fluid Dynamics in Belgium, and at Yale University, New Haven, CT.

Keynote 3:

Speaker: Dr. Rumi Ghosh, Robert Bosch LLC, CA

Bio:

Dr. Rumi Ghosh is a senior data mining engineer at Robert Bosch, LLC. Her areas of research include data mining, machine learning and complex networks analysis. She received her Bachelors and Masters Degrees in Mathematics and Computing from Indian Institute of Technology, Kharagpur, India in 2007 and Ph.D. in Computer Science from University of Southern California in 2012. Her doctoral dissertation dealt with understanding the interplay of topology and dynamics in network analysis. During her Ph.D. and post-doctoral research in HP Labs, she focused on devising algorithms for connected networks of people. When she joined Bosch, she forayed into internet of things. Her responsibilities at Bosch include development of algorithms for real data mining and machine learning problems for a wide spectrum of domains ranging from manufacturing to supply chain management to demand forecasting. She has 3 filed patents and has over 30 research papers in internationally reputed conferences and journals in Computer Science, Physics and Mathematics such as KDD, WWW, ICWSM, WSDM and Physical Review. She chaired workshops and symposiums in the field of data mining focusing on domains of advanced manufacturing and internet of things in conferences like ICDM and IEEE Big Data Conference. She has been in the program committee and reviewer for many international conferences and journals including IJCAI, KDD, TKDD, TKDE, ICDE, ICWSM and ACM Hypertext to name a few.

PANEL- Big Data Analytics for Advanced Manufacturing: Challenges and opportunities

Moderator: Dr. Sudarsan Rachuri, Federal Program Officer and Technology Manager, Advanced Manufacturing Office, Department of Energy

Panelists

Dr. Matteo Bellucci, GE Global Research
Mr. Matthew Jacobsen, Air Force Research Laboratory
Dr. Ram Sriram, NIST
Prof. Sankaran Mahadevan, Vanderbilt University
Dr. Sivaramakumar Gopalasundaram, Cognizant
Prof. Soundar Kumara, Penn State University
Dr. Valerie R. Coffman, Xeometry

BIOS

Mr. Matthew Jacobsen is a project manager in the Manufacturing and Industrial Technologies Division of the Air Force Research Laboratory, where he leads efforts in data and value stream management, process optimization, and cyber-physical vulnerabilities analysis. Mr. Jacobsen's current focus is concerned with the modernization of shop floor and supply chain IT capabilities, in order to address emerging issues in Big Data Analytics, Cloud Services, and Internet of Things (IoT) technology. To this end, Mr. Jacobsen is leading an internationally recognized cyberinfrastructure development program within the area of Integrated Computation Materials Science and Engineering (ICMSE). This cyberinfrastructure employs state-of the-art technologies to provide a complete suite of data management and machine integration capabilities to research and manufacturing organizations around the United States.

Dr. Valerie Coffman is a graduate of Johns Hopkins and received a PhD in Physics from Cornell where she wrote software for studying the fracture properties of materials. After graduation, she spent 5 years at the National Institute of Standards and Technology (NIST) writing open source software for materials science research. Valerie joined Xometry as Chief Technology Officer in 2014.

Dr. Sivaramakumar Gopalasundaram working as manager in Cognizant in the department of Data Analytics. Dr. Siva has a doctorate degree in Adaptive systems from Indian Institute of Science, Bangalore, India. He has worked in the area of Supply chain optimization, business forecasting and data analytics in various industrial sectors such as chemical processes, discrete manufacturing, health care, semiconductor, retail, automotive and communication.

Tutorials

TUTORIAL 1: Tutorial 1: Large Scale Text Mining – Techniques and Applications

Presenters:

Ronen Feldman, Professor
Information Systems Department, School of Business Administration,
Hebrew University Mount Scopus, Jerusalem, ISRAEL 91905
Tel: 972-(0)2-588-3084
Fax: 972-(0)-2-588-1341
Email: Ronen.Feldman@huji.ac.il

Ron Bekkerman, Assistant Professor
Department of Information and Knowledge Management,Faculty of Management
University of Haifa, Mount Carmel, Haifa, ISRAEL 34988
Tel: 972-(0)-4-664-7921
Fax: 972-(0)-4-824-9194
Email: ronb@univ.haifa.ac.il

Abstract:

The proliferation of documents available on the Web and on corporate intranets is driving a new wave of text mining research and applications. This massive scale of information is driving a new wave of text mining research and applications. Earlier research addressed extraction of information from relatively small collections of well-structured documents such as newswire or scientific publications. Text mining from the other corpora such as the web requires new techniques drawn from data mining, machine learning, NLP, and information retrieval. Text mining requires preprocessing document collections (text categorization, information extraction, term extraction), storage of the intermediate representations, analysis of these intermediate representations (distributional analysis e.g. word2vec, clustering, trend analysis, association rules, etc.), and visualization of the results. In this tutorial we will present the algorithms and methods used to build text mining systems. The tutorial will cover the state of the art in this rapidly growing area of research, including recent advances in unsupervised methods for extracting facts from text and methods used for web-scale mining. We will also present several real world applications of text mining. Special emphasis will be given to lessons learned from years of experience in developing real world text mining systems, including recent advances in sentiment analysis and information extraction and how to handle user generated text such as blogs and user reviews.

TUTORIAL 2: Trajectory Data Mining

Presenters:

Zhenhui (Jessie) Li,Assistant Profssor
Penn State University
Email: jessieli@ist.psu.edu

Fei Wu,PhD student
Penn State University
Email: fxw133@psu.edu

Jiawei Han,Professor
Univ. of Illinois at Urbana-Champaign
Email: hanj@cs.uiuc.edu

Abstract:

The advances in location-acquisition technologies and the prevalence of location-based services have generated massive spatial trajectory data, which represent the mobility of a diversity of moving objects, such as people, vehicles, and animals. Such trajectories offer us unprecedented information to understand moving objects and locations that

could benefit a broad range of applications in business, transportation, ecology, and many more. These important applications in turn call for novel computing technologies for discovering knowledge from trajectory data.

In this tutorial, we present a comprehensive, organized, and systematic survey on methodologies and algorithms on trajectory data mining. The tutorial will first give an overview of basic definitions, applications, data collection, data pre-processing, and patterns in the field of trajectory data mining. Then we will focus on three fundamental categories of trajectory patterns: (1) periodic pattern mining; (2) moving object relationship detection based on the spatial-temporal interactions which include friend relationship, follower/leader relationship, attraction/avoidance relationship, moving-together patterns, and clusters; and (3) semantic trajectory mining using external contexts. We will explore the connections, differences, and limitations of these existing techniques. Finally, we will discuss the use of trajectories in real-world applications such as recommendation, urban computing, and crime inference. We will conclude by discussing the exciting open topics in trajectory data mining.

Tutorial PPT

TUTORIAL 3: Large Scale Matrix Factorization

Presenters:

Fei Wang
Cornell University
Email:few2001@med.cornell.edu

Wei Tan
IBM T. J. Watson Research Center
Email:wtan@us.ibm.com

Abstract:

Matrix factorization has been a computational tool that aroused considerable interests in recent years in various analytics problems, such as clustering, collaborative filtering and topic modeling. With the arrival of the big data era, the volume and dimensionality of the data samples have increased a lot, which makes traditional batch-mode single core memory based matrix factorization methodologies not applicable and many large scale matrix factorization technologies have emerged. This tutorial will review various kinds of matrix factorization algorithms and their large scale implementation methodologies. We will also discuss about the current challenges and future directions.

TUTORIAL 4: Dynamic Big Data Processing in the Web of Things: Challenges, Opportunities and Success Stories

Presenters:

Ljiljana Stojanovic
Fraunhofer IOSB ,Germany

Nenad Stojanovic
Nissatech, Serbia
Email: nenad.stojanovic@nissatech.com

Abstract:

The Web of Things (WoT) is about involving real-world objects in the complex, Web-wide communication. WoT reuses and leverages readily available and widely popular Web protocols, standards and blueprints to make data and services offered by objects more accessible. However, WoT is generating an enormous amount of data (big data), e.g. 1 million connected devices all sending a sensor reading (e.g., temperature) every second to an IoT cloud means 86.4 billion messages per day (roughly 170 times more than all tweets posted globally that same day) and the most crucial issue is how to ensure an efficient (real-time) processing of this data, by knowing that the real-world objects generates very dynamic data streams. Indeed, the next wave of Big Data is Dynamic Big Data arising from new opportunities for

ubiquitous sensing and control of smallest details in engineered and natural systems, through multitudes of heterogeneous sensors and controllers instrumenting these systems, which inherently contain dynamics in their daily operation and require its proper management in order to increase the operational effectiveness and competitiveness. This tutorial tackles the intersection of these two very emerging areas, i.e. an efficient dynamic big data processing and management in the context of Web of Things

More particularly, processing data from real-world objects requires (big) data processing a) close to Things (local reaction: "moving" services to local data), b) close to Services (global reaction: moving data to global services) and c) the two-side interaction between these two levels. In other words, the challenge is to ensure that the local processing reflects the relevant part of the global context (services should be decomposed) and the global processing can react on the dynamicity of the data collected locally (services have to be dynamically changed). This processing & communication pattern can be found in many big data use cases, starting from wearables-driven well-being/fitness scenarios till the sensor-based proactive maintenance in the complex manufacturing scenarios.

Based on the ongoing work of authors, this tutorial explains the most important challenges for realizing dynamic data processing in WoT, the business opportunities derived from such a processing architecture and explains several success stories.

TUTORIAL 5: Anomalous and Significant Subgraph Detection in Attributed Networks

Presenters:

Feng Chen
University at Albany – SUNY
Email: fchen5@albany.edu

Petko Bogdanov
University at Albany – SUNY
Email: pbogdanov@albany.edu

Daniel B. Neill
Carnegie Mellon University
Email: neill@cs.cmu.edu

Ambuj K. Singh
University of California, Santa Barbara
Email: ambuj@cs.ucsb.edu

Abstract:

Detection of anomalous and significant subgraphs in attributed networks has applications in social networks, bioinformatics, disease surveillance and others. Different from vectors-space, single-vertex or whole graph versions, subgraph detection is often framed as a maximization of a score function over included node/edge attributes, where all connected or compact subgraphs are considered. Connectivity and compactness constraints ensure that subgraphs reflect changes due to localized in-network processes. The resulting problems are combinatorial in nature and, hence, require the design of efficient algorithms that scale to large real-world networks.

In this tutorial, we will present a comprehensive review of the state-of-the-art methods for anomalous and significant subgraphs detection. First, we will classify popular score functions and structure constraints commonly used in the literature. Then we will review methods for static (planar, complex, and heterogeneous) and dynamic networks. We will illustrate the basic theoretical and algorithmic ideas and discuss specific applications in all the above settings.

Tutorial PPT

Workshops

W1: Big Data for Cloud Operations Management (BDCOM)		
Time	**Title**	**Presenter/Author**
01:30 pm - 01:50 pm	**Bdcom Openning Remarks**	
01:50 pm - 02:10 pm	Open Big Data Infrastructures To Everyone	Konstantinos Tsakalozos, Cory Johns, Kevin Monroe, Pete Vandergiessens, Andrew Mcleod, And Antonio Rosales
02:10 pm - 02:30 pm	Intercloud Brokerages Based On Pls Method For Deploying Infrastructures For Big Data Analytics	Katsunori Miura
02:30 pm - 02:50 pm	Identifying Performance Bottlenecks In Hive: Use Of Processor Counters	Alexander Shulyak And Lizy John
02:50 pm - 03:10 pm	Data-Driven Cloud-Based It Services Performance Forecasting	Genady Grabarnik, Mauro Tortonesi, And Larisa Shwartz
03:10 pm - 03:30 pm	On-Demand Data Analytics In Hpc Environments At Leadership Computing Facilities: Challenges And Experiences	John Harney, Seung-Hwan Lim, Sreenivas Sukumar, Dale Stansberry, And Peter Xenopoulos
03:30 pm - 03:50 pm	**Coffee Break**	
03:50 pm - 04:10 pm	Leveraging Large Sensor Streams For Robust Cloud Control	Alok Singh, Eric Stephan, Todd Elsethagen, Matt Macduff, Bibi Raju, Malachi Schram, Kerstin Kleese Van Dam, Darren J Kerbyson, And Ilkay Altintas
04:10 pm - 04:30 pm	FINE-GRAINED POWER ANALYSIS OF EMERGING GRAPH PROCESSING WORKLOADS FOR CLOUD OPERATIONS MANAGEMENT	Shuang Song, Xinnian Zheng, Andreas Gerstlauer, And Lizy K. John
04:30 pm - 04:50 pm	MOTIVATING DYNAMIC FEATURES FOR RESPONSE TIME ESTIMATION WITHIN IT OPERATIONS MANAGEMENT	Kayhan Moharreri
04:50 pm - 05:10 pm	Holistic Disaster Recovery Approach For Big Data Nosql Workloads	Aharon Abadi, Ashraf Haib, Roie Melamed, Alaa Nassar, Aidan Shribman, And Hisham Yasin
05:10 pm - 05:30 pm	**Closing Remarks And Discussions**	

W2: 2nd International Workshop on Big Data for Sustainable Development		
Time	**Title**	**Presenter/Author**
8:00am-8:20am	Opening remarks	**Program Chair:** Aki-Hiro Sato, Kyoto **Program Co-Chairs:** Dr. Laura Irina Rusu Ms. Gandhi Sivakumar
8:20am-10:00am	A study for tourism policy making support based on log-data of Wi-Fi access points	Yu Ichifuji,, Noriaki Koide
	Estimation of National Tourism Statistics	Noriaki Koide
	Hotel Plan Popularity Factor Analysis of Hotels in the Keihanshin Region	Hiroshi Tsuda, Masakazu Ando, Yu Ichifuji
	Measuring Activities and Values of Industrial Clusters based on Job Opportunity Data Collected from an Internet Japanese Job Matching Site	Aki-Hiro Sato, Tsutomu Watanabe
10:00am-10:20am	**Coffee Break**	
10:20am-12:00am	Finding Effective Ways to Understand Sustainability Management of Best-in-class Financial Institutions in the Age of Big Text Data	Chu-hua Kuei, Ren-raw Chen,
	Peer-to-Peer Microlending Platforms: Characterization of Online Traits	Enrique Frias-Martinez, Gaurav Paruthi, Vanessa Frias-Martinez,
	Network Optimization of Food Flows in the U.S.	Caleb Robinson,Arezoo Shirazi, Mengmeng Liu, Bistra Dilkina,
	Multi-scalar Analysis of Geospatial Agricultural Data for Sustainability	Anne Denton, Mostofa Ahsan, David Franzen, John Nowatzki,
12:00pm-1:20pm	**Lunch Break**	
1:20pm-3:30pm	URBAN-NET: A Network-based Infrastructure Monitoring and Analysis System for Emergency Management and Public Safety	Sangkeun Lee, Liangzhe Chen, Sisi Duan, Supriya Chinthavali, Mallikarjun Shankar, B. Aditya Prakash
	Unravelling the Myth of Big Data and Artificial Intelligence in Sustainable Natural Resource Development	Gandhi Sivakumar, Drew Johnson, Rashida Hodge
	Short Break (2:10pm-2:15pm)	
	Solar Irradiance Forecasting by Machine Learning for Solar Car Races	Xiaoyan Shao, Siyuan Lu, Theodore G. van Kessel, Leda Daehler, Jeffrey Cwagenberg, Alan Li, Hendrik F. Hamann,
	Mixed Data and Classification of Transit Stops	Laura Tupper, David Matteson, John Handley
	Crowdsensing and Analyzing Micro-Event Tweets for Public Transportation Insights	Philips Kokoh Prasetyo, Thong Hoang, Pei Hua Cher, Ee-Peng Lim
3:30pm-3:50pm	**Coffee Break**	
3:50pm-4:40pm	Spatial-Crowd: A Big Data Framework for Efficient Data Visualization	Shahbaz Atta, Bilal Sadiq, Akhlaq Ahmad, Sheikh Nasir Saeed, Emad Felemban
	A framework for evaluating urban land use mix from crowd-sourcing data	Luciano Gervasoni, Mart íBosch Padros, Serge Fenet
4:40pm-4:45pm	**Short Break**	
4:45pm-5:25pm	**Invited Talk** Tourism, the Experience economy and Walkable Urban Places	Christopher B. Leinberger
5:25pm-5:30pm	Closing Remarks	**Program Chair:** Aki-Hiro Sato, Kyoto **Program Co-Chairs:** Dr. Laura Irina Rusu Ms. Gandhi Sivakumar,

W3: 3rd Workshop on Advances in Software and Hardware for Big Data to Knowledge Discovery (ASH) & 4th International Workshop on Distributed Storage Systems and Coding for Big Data (DSSCB)		
Time	**Title**	**Presenter/Author**
8:15am – 8:30am	**ASH & DSSCB opening remarks**	
8:30am – 8:55am	A Scalable and Composable Map-Reduce System	Mahwish Arif, Hans Vandierendonck, Dimitrios S. Nikolopoulos, and Bronis R. de Supinski
8:55am – 9:20am	Big Data Analytics on HPC Architectures: Performance and Cost	Peter Xenopoulos, Jamison Daniel, Michael Matheson, and Sreenivas Sukumar
9:20am – 9: 45am	Evaluation of K-Means Data Clustering Algorithm on Intel Xeon Phi	Sunwoo Lee, Wei-keng Liao, Ankit Agrawal, Nikos Hardavellas, and Alok Choudhary
9:45am – 10:00am	Building a Research Data Science Platform from Industrial Machines	Fang Liu and Duen Horng Chau
10:00am – 10:20am	**Coffee Break**	
10:20am – 10:45am	Visually Programming Dataflows for Distributed Data Analytics	Lauritz Thamsen, Thomas Renner, Marvin Byfeld, Markus Paeschke, Daniel Schröder, and Felix Böhm
10:45am – 11:05am	A Geohydrologic Data Visualization Framework with an Extendable User Interface Design	Yanfu Zhou, Jieting Wu, Lina Yu, Hongfeng Yu, and Zhenghong Tang
11:05am-11:30am	Efficient Portfolio Allocation with Sparse Volatility Estimation for High-Frequency Financial Data	Jian Zou and Chuqin Huang
11:30am – 11:55am	Accelerating Mathematical Knot Simulations with R on theWeb	Hui Zhang, Juan Lin, Di Zhong, and Yiwen Zhong
noon – 1:00pm	**Lunch Break**	
1:00pm – 1:25pm	A Workload Aware Model of Computational Resource Selection for Big Data Applications	Amit Gupta, Weijia Xu, Natalia Ruiz-Juri, and Kenneth Perrine
1:25pm – 1:50pm	Supporting Large Scale Connected Vehicle Data Analysis using Hive	Weijia Xu, Natalia Ruiz-Juri, Amit Gupta, Amanda Deering, Chandra Bhat, James Kuhr, and Jackson Archer
1:50pm – 2:15pm	Legion-based Scientific Data Analytics on Heterogeneous Processors	Lina Yu and Hongfeng Yu
2:15pm – 2:40pm	Materials Discovery: Understanding Polycrystals from Large-Scale Electron Patterns	Ruoqian Liu, Ankit Agrawal, Wei-keng Liao, Marc De Graef, and Alok Choudhary
2:45pm – 3:30pm	**Reserved for Panel/Discussion**	
3:30pm - 3:50pm	**Coffee Break**	
3:50pm - 4:15pm	Towards Optimizing Large-Scale Data Transfers with End-to-End Integrity Verification	Si Liu, Eun-Sung Jun, Rajkumar Kettimuthu, Xian-He Sun, and Michael Papka
4:15pm - 4:40pm	EStore: An Effective Optimized Data Placement Structure for Hive	Xin Li, Hui Li, Zhihao Huang, Bing Zhu, and Jiawei Cai
4:40pm - 5:05pm	SS-Dedup: A High Throughput Stateful Data Routing Algorithm for Cluster Deduplication System	Zhihao Huang, Hui Li, Xin Li, and Wei He
5:05pm - 5:30pm	CoLoc: Distributed Data and Container Colocation for Data-Intensive Applications	Thomas Renner, Lauritz Thamsen, and Odej Kao
5:30pm - 5:55pm	Persisting In-Memory Databases Using SCM	Ellis Giles, Kshitij Doshi, and Peter Varman

W4: Open Science in Big Data (OSBD)		
Time	**Title**	**Presenter/Author**
1:30pm – 1:40pm	**Introduction**	
1:40pm - 2:30pm	Big neuroimaging and open science Keynote Speaker: HHMI Janelia Farms	Jeremy Freeman
2:30pm - 2:50pm	Public health surveillance Invited Speaker: Oak Ridge National Laboratory	Arvind Ramanathan
2:50pm - 3:10pm	Open source big data in industry Invited Speaker: Lucidworks	Jake Mannix
3:10pm - 3:30pm	Invited Speaker: Georgia Institute of Technology	Ling Liu
3:30pm - 3:50pm	**Coffee Break**	
3:50pm – 4:10pm	"Dask & Numba: Simple Libraries for Optimizing Scientific Python Code" Invited Speaker: Continuum Analytics	Jim Crist
4:10pm - 4:30pm	"Promoting Open Science in the University" Invited Speaker: University of Washington eScience Institute	Jake VanderPlas
4:30pm – 4:40pm	**Panel Session: Open Science in Government, Academia, and Industry** **Panel members: Arvind Ramanathan, Jake Mannix, Ling Liu, Jim Crist, Jake VanderPlas**	
4:40pm – 4:55pm	Making Massive Computational Experiments Painless	Hatef Monajemi, David Donoho, and Victoria Stodden
4:55pm – 5:10pm	Too Big to Mail: On the Way to Publish Large-scale Mobile Analytics Data	Ella Peltonen, Eemil Lagerspetz, Petteri Nurmi, and Sasu Tarkoma
5:10pm - 5:25pm	**Coffee Break**	
5:25pm – 5:40pm	Content-based Recommendation for Podcast Audio-items using Natural Language Processing Techniques	Zhou Xing, Marzieh Parandehgheibi, Fei Xiao, Nilesh Kulkarni, and Chris Pouliot
5:40pm – 5:55pm	PinterNet: A Thematic Label Curation Tool for Large Image Datasets	Ruoqian Liu, Diana Palsetia, Arindam Paul, Reda Al-Bahrani, Dipendra Jha, Wei-keng Liao, Ankit Agrawal, and Alok Choudhary
5:55pm – 6:10pm	A Big Data Platform Integrating Compressed Linear Algebra with Columnar Databases	Vishnu Gowda Harish, Vinay Kumar Bingi, and John A Miller
6:10pm – 6:25pm	Implementing Dictionary Learning in Apache Flink, Or: How I Learned to Relax and Love Iterations	Geoffrey Mon, Milad Makkie, Xiang Li, Tianming Liu, and Shannon Quinn
6:25pm – 6:30pm	**Wrap-up and Concluding Remarks**	

W5: Workshop on Real-time and Stream Analytics in Big Data		
Time	**Title**	**Presenter/Author**
8:00am – 8.25am	**Introduction**	
8:25am - 8:50am	Implementing Trajectory Data Stream Analysis in Parallel	Yongyi Xian, Chuanfei Xu, and Yan Liu
8:50am - 9:15am	A Glue Language for Event Stream Processing	Sylvain Hallé, Sébastien Gaboury, and Raphaël Khoury
9:15am – 9:40am	An FPGA-Based Low-Latency Network Processing for Spark Streaming	Kohei Nakamura, Ami Hayashi, and Hiroki Matsutani
10:00am - 10:20am	**Coffee Break**	
10:20am – 10:45am	A multi-layer software architecture framework for adaptive real-time analytics	Athena Vakali, Paschalis Korosoglou, and Pavlos Daoglou
10:45am – 11:10am	Predicting the Shape and Peak Time of News Article Views	Yaser Keneshloo, Shuguang Wang, Eui-Hong Han, and Naren Ramakrishnan
10:10am - 11:35am	Real-time processing of proteomics data	Christopher Hillman, Andrew Cobley, karen Petrie, and Mark Whitehorn
11:35am - 12:00am	Handling Delayed Labels in Temporally Evolving Data Streams	Joshua Plasse and Niall Adams

W6: Application of Big Data for Computational Social Science		
Time	**Title**	**Presenter/Author**
11:05am-11:10 am	Opening	Akira Ishii
11:10am – 12:10pm	**Session1 Moral and Politics**	
15 mins for each	Tweet Sentiment as Proxy for Political Campaign Momentum	K.M. George, Zenia Arora,
	Pricing the Woman Card: Gender Politics between Hillary Clinton and Donald Trump	Katerina Doka, Mingqiang Xue, Dimitrios Tsoumakos, Panagiotis Karras, Alfredo Cuzzocrea, and Nectarios Koziris
	Quantifying moral foundations from various topics on Twitter **conversations**	Rishemjit Kaur, Kazutoshi Sasahara,
Lunch Time(12:10-14:00)		

14:00pm – 15:45pm	**Session2 SocialMedia and Web1**	
15 mins for each	Language independent Big-Data system for the prediction of user location on Twitter	Jaime Alonso-Lorenzo, Enrique Costa-Montenegro, Milagros Fernández-Gavilanes,
	Automated Classification of ISIS Twitter Accounts Using Content-Based and Network-Based Features	Daniel Xie, Jiejun Xu, Tsai-Ching Lu
	Uncovering Information Flow Among Users by Time-Series Retweet Data: who is a friend of whom on Twitter?	Yuka Kamiko, Mitsuo Yoshida, Hirotada Ohashi, Fujio Toriumi
	Analytical method of web user behavior using Hidden Markov Model	Hirotaka Kawazu, Masanori Takano, Kazuya Wada, Ichiro Fukuda
	User-generated Content Curation with Deep Convolutional Neural Networks	Ruben Tous, Otto Wust, Mauro Gomez,Jonatan Poveda, Marc Elena, Jordi Torres, Barcelona Mouna Makni, Eduard Ayguadé,
	Finding Informative Comments for Video Viewing	Seungwoo Choi, Aviv Segev,
	Prediction of Information Diffusion in Social Networks using Dynamic Carrying Capacity	Anahita Davoudi, Mainak Chatterjee,
	Classifying Twitter User Judgments of Rumors Using Distributed Representations of Words	Armineh Nourbakhsh, Xiaomo Liu, Sameena Shah, Rui Fang, Quanzhi Li
Coffe Break (15:45-16:25)		
16:25pm – 17:55pm	**Session3 Marketing**	
15mins for each	Forecasting Nike's Sales using Facebook Data	Linda Camilla Boldt, Vinothan Vinayagamoorthy, Florian Winder, Melanie Schnittger, Mats Ekran, Raghava Rao Mukkamala, Niels Buus Lassen, Benjamin Flesch, Abid Hussain, Ravi Vatrapu,
	Automated Classification of ISIS Twitter Accounts Using Content-Based and Network-Based Features	Daniel Xie, Jiejun Xu, Tsai-Ching Lu
	Uncovering Information Flow Among Users by Time-Series Retweet Data: who is a friend of whom on Twitter?	Yuka Kamiko, Mitsuo Yoshida, Hirotada Ohashi, Fujio Toriumi,
	Application of Integer-Valued Autoregressive Model to Hit Phenomena	Yasuko Kawahata, Tamio Koyama
	Nowcast of firms' sales using POS data toward the stability of stock market	Atushi Ishikawa, Shouji Fujimoto, Takayuki Mizuno
	A New Approach to Building the Interindustry Input--Output Table Using Block Estimation Techniques	Ryohei Hisano
	Leveraging Social Big Data for Performance Evaluation of E-Commerce Websites	Eyad Makki, Lin-Ching Chang,
	When Do Luxury Cars Hit the Road? Findings by A Big Data Approach	Yang Feng Jiebo Luo,
	Experimentation and the Diffusion of Technology in China: Using Big Data to explore Consumer Channel Choice	Ashley Lloyd, Mario Antonioletti, Terence Sloan,

W7: Methods To Manage Heterogeneous Big Data And Polystore Databases		
Time	**Talk**	**Speaker(S)**
13:30-13:45	Welcome	Vijay Gadepally, Michael Stonebraker
13:45-14:15	Keynote Address (Title Tbd)	Fatma Ozcan, Ibm Research
14:20-14:40	Towards A Heterogeneous, Polystore-Like Data Architecture For The Us Department Of Veteran Affairs (Va) Enterprise Analytics	Edmon Begoli,Jack Bates,Derek Kistler
14:45-15:05	Benchmarking Polystores: The Cloudmdsql Experience	Boyan Kolev,Raquel Pau, Patrick Valduriez, Ricardo Jimenez, Jose Pereira
15:10-15:30	Analytics-Driven Data Ingestion And Derivation In The Awesome Polystore	Subhasis Dasgupta, Kevin Coakley, Amarnath Gupta
15:30-15:50	A Semantic Approach To Polystores	Evgeny Kharlamov, Konstantina Mpereta, Dimitris Bilidas,Ernesto Jimenez-Ruiz, Steffen Lamparter, Christian Neuenstadt, Oezguer Oezcep, Ahmet Soylu,Christoforos Svingos,Guohui Xiao, Dmitriy Zheleznyakov,Diego Calvanese, Ian Horrocks, Martin Giese, Yannis Ioannidis, Yannis Kotidis, Ralf Moeller, Arild Waaler
15:55-16:25	**Break**	
16:25-17:05	Keynote Address	Luna Dong
17:10-17:20	Digree: A Middleware For A Graph Databases Polystore	Vasilis Spyropoulos, Christina Vasilakopoulou, Yannis Kotidis
17:25-17:45	Hobbits: Hadoop And Hive Based Internet Traffic Analysis	Abdeltawab Hendawi,
17:50-18:20	Discussion: What Are Polystores, Multistores,...	Tim Mattson, Dave Maier,
18:20-18:30	Discussion, Closing	Tim Mattson, Vijay Gadepally, Micheal Stonebraker

W9: 1st IEEE International Workshop on Big Spatial Data		
Time	**Title**	**Presenter/Author**
7:30am – 8.00am	**Registration**	
8:00am - 8:10am	**Welcome**	
	Paper Presentation Session: Algorithms and Data Quality	

8:10am-8:40am	Big Data Computation of Taxi Movement in New York City	Joya Deri, Franz Franchetti, and Jos é M.F. Moura
8:40am - 9:10am	A Comparative Study of Dual-tree Algorithm for Computing 2-Body Statistics in Spatial Data	Chengcheng Mou, Shaoping Chen, and Yi-Cheng Tu
9:10am – 9:40am	The SMART Approach to Comprehensive Quality Assessment of Site-Based Spatial-Temporal Data	Douglas Galarus and Rafal Angryk
9:40am -10:00am	Towards a Provenance-Aware Spatial-Temporal Architectural Framework for Massive Data Integration and Analysis (Short Paper)	Ivens Portugal, Paulo Alencar, and Donald Cowan
10:00am - 10:20am	**Coffee Break**	
	Paper Presentation Session: Platforms and Applications	
10:20am – 10:40am	Big Data Development Platform For Engineering Applications (Short Paper)	Chien-Heng Wu, Whey-Fone Tsai, Franco Lin, Wen-Yi Chang, Shi-Ching Lin, and Chao-Tung Yang
10:40am – 11:10am	A Survey of the Big Spatial Data Technology Landscape	Andrew Hulbert, Anthony Fox, James Hughes, Thomas Kunicki, Matthew Zimmerman, and Christopher Eichelberger
10:10am - 11:30am	IBM PAIRS Curated Big Data Service for Accelerated Geospatial Data Analytics and Discovery (Short Paper)	Siyuan Lu, Xiaoyan Shao, Marcus Freitag, Levente Klein, Jason Renwick, Fernando Marianno, Conrad Albrecht, and Hendrik F. Hamann
11:30am - 12:00pm	Linked Data View Methodology and Application to BIM Alignment and Interoperability	Holly Ferguson and Charles Vardeman
12:00pm -2:00pm	**Lunch**	
2:00 pm – 3:30 pm	**Paper Presentation Session: Clustering**	
2:00 pm-2:30 pm	Using Parallel Hierarchal Clustering to Address Spatial Big Data Challenges	Alan Woodley, Shlomo Geva, Richi Nayak, Ling-Xiang Tang, and Timothy Chappell
2:30pm -3:00 pm	Adapting K-Means Clustering to identify Spatial Patterns in Storms	Upa Gupta, Kulsawasd Jitkajornwanich, Ramez Elmasri, and Leonidas Fegaras
3:00 pm -3:30pm	Symmetric Repositioning of Bisecting K-means Centers for Increased Reduction of Distance Calculations for Big Data Clustering	Yu Zhuang
3:30pm -3:50pm	**Coffee Break**	
3:50 pm – 4:45 pm	**Keynote: Big Spatial Data at Facebook** **(Xiaoming Gao, Research Scientist at Facebook Inc.; Saurav Mohapatra, Software Engineer at Facebook Inc.; Lihan Bin, Software Engineer at Facebook Inc.)**	
	Paper Presentation Session: Imagery Analysis	
4:45pm – 5:15pm	Determining Feature Extractors for Unsupervised Learning on Satellite Images	Behnam Hedayatnia, Mehrdad Yazdani, Mai Nguyen, Jessica Block, and Ilkay Altintas
5:15pm – 5:45 pm	Large-Scale Solar Panel Mapping from Aerial Images Using Deep Convolutional Networks	Jiangye Yuan, Hsiu-Han Yang, Olufemi Omitaomu, and Budhendra Bhaduri
3:30pm -3:50pm	**Adjourn**	

W10: Workshop on Big Data and Machine Learning in Telecom		
Time	**Title**	**Presenter/Author**
8:00am – 8.05am	**Introduction**	
8:05am - 8:50am	Keynote Speech: Big Data Analytics in Mobile Environments	Hui Xiong
8:50am - 9:35am	Keynote Speech	Ye Ouyang
9:35am - 10:00am	Preliminary Big Data in a 5G Test Network	Teemu Kanstrén, Jussi Liikka, Jukka Mäkelä, Markus Luoto, and Jarmo Prokkola
10:00am - 10:20am	**Coffee Break**	
10:20am - 10:45am	Quick Model Fitting Using a Classifying Engine	Yiming Kong, Hui Zang, and Xiaoli Ma
10:45am - 11:10am	WHAT: A Big Data Approach for Accounting of Modern Web Services	Martino Trevisan, Idilio Drago, Marco Mellia, Han Hee Song, and Mario Baldi
11:10am - 11:35am	Spark-based Rare Association Rule Mining for Big Datasets	Ruilin Liu, Kai Yang, Yanjia Sun, Tao Quan, and Jin Yang
11:35am – 12.00pm	Evaluating Machine Learning Algorithms for Anomaly Detection in Clouds	Anton Gulenko, Marcel Wallschläger, Florian Schmidt, Odej Kao, and Feng Liu

W11: 4th Workshop on Scalable Cloud Data Management		
Time	**Title**	**Presenter/Author**
8:00-8:05pm	Opening Remarks	Norbert Ritter, Felix Gessert
8:05am-8:50am	Keynote Address: A Storage Perspective on Scalable Data Management in the Cloud: Thinking Beyond the Present	Sangeetha Seshadri (IBM Almaden Research Center)
8:50am-10:35am	**Session I: Data Management**	
	NoSQL Schema Evolution and Big Data Migration at Scale	Meike Klettke (University of Rostock, Germany)
	Analyzing the Performance of Data Replication and Data Partitioning in the Cloud: the Beowulf Approacht	Alexander Stiemer (University of Basel, Switzerland)
	Is Elasticity of Scalable Databases a Myth?	Daniel Seybold (Ulm University, Germany)
10:35am-10:50am	**Coffee Break**	
10:50am-12:35pm	**Session II: Cloud Databases and Systems**	
	Non-deep CNN for Multi-Modal Image Classification and Feature Learning: An Azure-based Model	Sohini Roychowdhury (University of Washington, USA)
	Understanding performance interference in multi-tenant cloud databases and web applications	Miguel Xavier (PUCRS, Brazil)
	Container-Based Virtualization for Byte-Addressable NVM Data Storage	Ellis Giles (Rice University, USA)
12:35pm-1:35pm	**Lunch Break**	
1:35pm-3:20pm	**Session III: Big Data**	
	Towards An Integrated Health Research Process: A Cloud-based Approach	Matthieu-P. Schapranow (Hasso Plattner Institute, Germany)
	BINARY: A Framework for Big Data Integration for Ad-hoc Querying	Farhana Zulkernine (Queen's University, Canada)
	Scheduling Big Data Workflows in the Cloud under Budget Constraints	Aravind Mohan (Wayne State University, USA)
3:20pm-3:40pm	**Coffee break**	
3:40-4:50pm	**Session IV: Big Data**	
	Big data availability: Selective partial checkpointing for in-memory database queries	Daniel Playfair (SAP, United Kingdom)
	Model-driven Deployment and Management of Workflows on Analytics Frameworks	Merlijn Sebrechts (Ghent University, Belgium)
4:50-5:40pm	**Special Session: Smart Data**	
	The Smart Data Program	Nico Roedder (FZI Research Center for Information Technology, Germany)
	The digital transformation and smart data analytics: An overview of enabling developments and application areas	

W12: 2nd International Workshop on Methodologies to Improve Big Data Projects (MIDP-2016)		
Time	**Title**	**Presenter/Author**
1:30pm-1:40 pm	Introduction / Opening Remarks	Jeffrey Saltz
1:40 pm – 3:20 pm (25mins for each)	Software Engineering for Big Data Projects: Domains, Methodologies and Gaps	Vijay Dipti Kumar & Paulo Alencar
	Big Data Team Process Methodologies: A Literature Review and the Identification of Key Factors for a Project's Success	Ivan Shamshurin
	Bad Big Data Science	Frank Haug
	Not All Software Engineers Can Become Good Data Engineers	Sibel Yilmazel & Ozgur Yilmazel
3:20 pm – 3:50pm	**Break**	
3:50 pm:5:30 pm (25mins for each)	Mapping the old data mining process model CRISP-DM into the NIST big data reference architecture	Nancy Grady, invited speaker
	Progression Analysis of Signals: Extending CRISP-DM to Stream Analytics	Pankush Kalgotra & Ramesh Sharda
	Evaluation-Driven Research in Data Science: Leveraging Cross-Field Methodologies	Bonnie Dorr
	A Hacking Toolset for Big Tabular Files	Toshiyuki Shimono

W13: Big Data Challenges, Research, and Technologies in the Earth and Planetary Sciences		
Time	**Title**	**Presenter/Author**
8:00am – 8.05am	**Introduction**	
8:05am - 8:25am	Using Cloud Bursting to Count Trees and Shrubs in Sub-Saharan Africa	Michael Requa, Garrison Vaughan, John David, and Ben Cotton
8:25am - 8:45am	A New Parallel Python Tool for the Standardization of Earth System Model Data	Kevin Paul, Sheri Mickelson, and John Dennis
8:45am - 9:05am	Three-Dimensional Spatial Join Count exploiting CPU Optimized STR-R-Tree	Takahiro Nishimichi, Ryuya Mitsuhashi, Hideyuki Kawashima, and Osamu Tatebe
9:05am - 9:25am	SciSpark: Highly Interactive In-Memory Science Data Analytics	Brian Wilson, Rahul Palamuttam, Kim Whitehall, Chris Mattmann, Alex Goodman, Maziyar Boustani, Sujen Shah, Paul Zimdars, and Paul Ramirez
9:25am – 9.45am	Visualization and Diagnosis of Earth Science Data through Hadoop and Spark	Zhou, S., X. Li, T. Matsui, and W.-K. Tao
9:45am – 10.05am	Modeling Martian Thermal Inertia in a Distributed Memory High Performance Computing Environment	Jason Laura and Robin Fergason
10:05am - 10:20am	**Coffee Break**	
10:20am - 10:40am	Implementing Connected Component Labeling as a User Defined Operator for SciDB	Amidu Oloso, Kwo-Sen Kuo, Thomas Clune, Paul Brown, Alex Poliakov, Hongfeng Yu
10:40am - 11:00am	Distributed and cloud-based multi-model analytics experiments on large volumes of climate change data in the Earth System Grid Federation (ESGF) eco-system	S. Fiore, M. Płóciennik, C. Doutriaux, C. Palazzo, J. Boutte, T. Żok, D. Elia, M. Owsiak, A. D'Anca, Z. Shaheen, R. Bruno, M. Fargetta, M. Caballer, G. Moltó, I. Blanquer, R. Barbera, M. David, G. Donvito, D. N. Williams, V. Anantharaj, D. Salomoni, and G. Aloisio
11:00am - 11:20am	Where Big Data meets Linked Data: Applying standard data models to environmental data streams	Adam Leadbetter, Adam Shepherd, Damian Smyth, Robert Fuller, and Eoin O'Grady
11:20am – 11:30am	**Open Discussion on Papers**	
11:30am - 11:50am	**Community Reports - George Percival – Open Geospatial Consortium**	
11:50am - 12:10pm	**Community Reports – Ben Evans – National Computational Infrastructure, Australia**	
12:10pm - 12:30pm	**Community Reports – Mike Little - NASA**	

W15: IEEE Workshop on Big Data Metadata and Management (BDMM '16)		
Time	**Title**	**Presenter/Author**
9:00am-9:20am	Opening Remark: IEEE Big Data Initiative (BDI)	Mahmoud Daneshmand
9:20am-10:00am	**Session I: Metadata Management**	
	MetaStore: Metadata Framework for Scientific Data Repository	Ajinkya Prabhune, Anil Keshav, Hasebullah Ansari, Rainer Stotzka Michael Gertz, Juergen Hesser
	Fault-tolerant Data Transfer Strategy Using Bandwidth Scheduling Service in High-performance Networks	Liudong Zuo, Michelle Zhu
10:00am-10:20am	**Coffee Break**	
10:20am-11:20am	**Keynote**	
	Standards for Big Datasets	Robby Robson
11:20am-12:00pm	**Session II: Computational Management**	
	Facilitating Reproducible Research by investigating Computational metadata	Priyaa Thavasimani, Paolo Missier
	Automated Schema Extraction for PID Information Types	Ulrich Schwardmann
12:00pm-2:00pm	**Lunch Break**	
2:00pm-3:30pm	**Session III: Metadata and Applications**	
	Detecting Spammers on Social Networks Based on a Hybrid Model	Guangxia Xu, *et. al.*
	Linked Data Platform for Building Resilience Based Applications and Connecting API Access Points with Data Discovery Techniques	Holly Ferguson
	Constellation: A Science Graph Network for Scalable Data and Knowledge Discovery in Extreme-Scale Scientific Collaborations	Sudharshan S. Vazhkudai, *et. al.*
3:30pm-3:50pm	**Coffee break**	
3:50pm-4:50pm	**Panel Discussion**	
	Challenges and Opportunities in Standardizing Big Data Management	Moderator: Alex Kuo Panelists: Robby Robson, Cherry Tom, Mahmoud Daneshmand, Kathy Grise, Yinglong Xia, Dave Belanger
4:50pm-5:00pm	Close Remark	Yinglong Xia

W16: BDSG: Big Data in Smart Grids		
Time	**Title**	**Presenter/Author**
7:30am – 8.00am	**Registration**	
8:00am-8:20am	**Introduction:** Big Data Enablers for Open Access Smart Grids (OASIS)	Manuel Rodriguez, BDSG Co-chair, UPRM
8:20am – 9:20am	**Keynote: TBD**	Vipin Chaudhary Program Director, CISE/ACI National Science Foundation
9:20am - 9:40am	Leveraging User Expertise in Collaborative Systems for Annotating Energy Datasets	Hông-Ân Cao, Felix Rauchenstein, Tri Kurniawan Wijaya, Karl Aberer, and Nuno Nunes,
9:40am – 10:00am	Leveraging Cloud Computing to Convert the Non-Intrusive Load Monitor into a Powerful Framework for Grid-Responsive Buildings	Robert Cox, Saman Mostafavi, John Troxler, and Benjamin Futrell
10:00am - 10:20am	**Coffee Break**	
10:20am - 10:40am	Big Data, Better Energy Management and Control Decisions for Distribution Systems in Smart Grid	Shady Khalil, Haitham Abu-Rub, and Amira Mohamed
10:40am - 11:00am	Detecting Non-Technical Energy Losses through Structural Periodic Patterns in AMI data	Marina Papatriantafilou, Magnus Almgren, Vincenzo Gulisano, Olaf Landsiedel, Joris van Rooij, and Viktor Botev
11:00am - 11:20am	Temporal Association Rules For Electrical Activity Detection in Residential Homes	Hông-Ân Cao, Tri Kurniawan Wijaya, Karl Aberer, and Nuno Nunes
11:20am-11:40am	Investigation of forecasting methods for the hourly spot price of the Day-Ahead Electric Power Markets	Radha krishnan Angamuthu and Prakash Ranganathan
11:40am-12:00pm	Lossless Compression of High-Frequency Voltage and Current Data in Smart Grids	Andreas Unterweger and Dominik Enge

W17: Solar & Stellar Astronomy Big Data (SABiD)		
Time	**Title**	**Presenter/Author**
13:30-13:45	**Introductions**	
13:45-14:10	Processing and Managing the Kepler Mission's Treasure Trove of Stellar and Exoplanet Data	Jon Jenkins
13:10-14:35	An Input Catalog and Target Selection for the Transiting Exoplanet Survey Satellite	Ryan Oelkers, Keivan Stassun, Joshua Pepper, Nathan De Lee, and Martin Paegert
14:35-15:00	Stream Multidimensional D^2 Phase Dispersion Statistic: Test Cases and Application to a Massive ”In Silico” Dataset	Nigul Olspert, Maarit Käpylä, and Jaan Pelt
15:00-15:25	Running Scientific Algorithms as Array Database Operators: Bringing the Processing Power to the Data	Simon Marcin and Andr é Csillaghy
15:30-15:50	**Coffee Break**	

15:50-16:15	A Data-Driven Analysis of Interplanetary Coronal Mass Ejecta and Magnetic Flux Ropes	Ruizhe Ma, Rafal Angryk, and Pete Riley
16:15-16:40	The Best of Both Worlds: Using Automatic Detection and Limited Human Supervision to Create a Homogenous Magnetic Catalog Spanning Four Solar Cycles	Amdres Muñoz-Jaramillo, Zachary Werginz, Juan Pablo Vargas-Acosta, Michael DeLuca, John Windmueller, Jie Zhang, Dana Longcope, Derek Lamb, Craig DeForest, Santiago Vargas-Dominguez, John Harvey, and Petrus Martens
16:40-16:55	Describing Solar Images with Sparse Coding for Similarity Search	Dustin Kempton, Michael Schuh, and Rafal Angryk
16:55-17:20	Spatiotemporal Interpolation for Solar Events from Mixed Data Sources	Soukaina Filali Boubrahimi, Berkay Aydin, Dustin Kempton, and Rafal Angryk
17:20-17:45	Indexing Spatiotemporal Relations in Solar Event Datasets	Berkay Aydin, Ahmet Kucuk, and Rafal Angryk
17:45-18:30	**Invited Talk**	Jack Ireland, NASA
18:30-18:40	Closing Discussion	

W18: Computational Archival Science		
Time	**Title**	**Presenter/Author**
8:45am-9:00am	Welcome	Mark Hedges (KCL)
9:00am-9:45am	Keynote	Mark Conrad (NARA)
9:45am-10:45am	**Paper Session I**	
	Exploring archives with probabilistic models: Topic Modelling for the valorisation of digitised archives of the European Commission	Simon Hengchen, Mathias Coeckelbergs, Seth van Hooland,Ruben Verborgh, Thomas Steiner
	Traces through time: A probabilistic approach to connected archival data	Sonia Ranade
	Opening Up Dark Digital Archives Through The Use of Analytics to Identify Sensitive Content	Jason R. Baron, Bennett B. Borden
10:45am-11:05am	**Coffee Break**	
11:05am-12:45pm	**Paper Session II**	
	Computational Provenance in DataONE: Implications for Cultural Heritage Institutions	Robert J. Sandusky
	Content-based Comparison for Collections Identification	Weijia Xu, Ruizhu Huang, Maria Esteva, Jawon Song, Ramona Walls
	Breaking Down the Invisible Wall to Enrich Archival Science and Practice	Kenneth Thibodeau
	Mind the explanatory gap: Quality from Quantity	Jenny Bunn
	Understanding Computational Web Archives Research Methods Using Research Objects	Emily Maemura, Christoph Becker, Ian Milligan
12:45pm-2:00pm	**Lunch Break**	
2:00pm-2:40pm	**Paper Session III**	
	Appraising Digital Archives with Archivematica	Michael Shallcross
	Mining and Analysing One Billion Requests to Linguistic Services	Marco Büchler, Greta Franzini, Emily Franzini, Thomas Eckart
2:40pm-3:30pm	**Panel Session: The future for research and education in CAS**	
3:30pm-4:00pm	**Coffee Break**	
4:00pm-5:30pm	**Group discussions and reporting back**	

W19: Third International Workshop on High Performance Big Graph Data Management, Analysis, and Mining (BigGraphs 2016)		
Time	**Title**	**Presenter/Author**
8:00am - 8:10am	Opening remarks	Mohammad Al Hasan
8:10am - 8:25am	On the Hyperbolicity of Large-Scale Networks	Iraj Saniee, Sean Kennedy, and Onuttom Narayan
8:25am - 8:40am	Parallel Graph Mining with Dynamic Load Balancing	Nilothpal Talukder and Mohammed Zaki
8:40am - 8:55am	Distributed Exact Subgraph Matching in Small Diameter Dynamic Graphs	Charith Wickramaarachchi, Rajgopal Kannan, Charalampos Chelmis, and Viktor Prasanna
8:55am - 9:00am	(optional) 5-minute break	
9:00am - 10:00am	Keynote talk	
10:00am - 10:20am	Coffee break	
10:20am - 10:35am	GraphFlow: Workflow-based Big Graph Processing	Sara Riazi and Boyana Norris
10:35am - 10:50am	Massive Graph Processing on Nanocomputers	Bryan Rainey and David Gleich
10:50am - 11:05am	GFP-X: A Parallel Approach To Massive Graph Comparison Using Spark	Stephen Bonner, John Brennan, Georgios Theodoropoulos, Ibad Kureshi, and Stephen McGough
11:05am - 11:20am	Deep Topology Classification: A New Approach for Massive Graph Classification	Stephen Bonner, John Brennan, Georgios Theodoropoulos, Ibad Kureshi, and Stephen McGough
11:20am - 11:25am	(optional) 5-minute break	
11:25am - 11:40am	Fast Reachability Computation on Big Attributed Graphs	Ka Wai Yung and Shi-Kuo Chang
11:40am - 11:55am	Fast distributed k-nn graph update	Thibault Debatty, Fabio Pulvirenti, Pietro Michiardi, and Wim Mees
11:55am - 12:10pm	An Incremental Local-First Community Detection Method for Dynamic Graphs	Hiroki Kanezashi and Toyotaro Suzumura

W20: 3rd Big Data Analytic Technology for Bioinformatics and Health Informatics Workshop (KDDBHI 2016)		
Time	**Title**	**Presenter/Author**
8:00am – 8:05am	**Chairs' Message, Donghui Wu, Xin Deng**	
8:05am - 8:25am	Drug Target Path Discovery on Semantic Biomedical Big Data	Fang Du, Shi Yingjie, Ting Li, Lijuan Song, and Xiaojun Gu
8:25am - 8:45am	Application of Big Data Analytics for Automated Estimation of CT Image Quality	Maitham Naeemi, Johnny Ren, Nathan Hollcroft, Adam Alessio, and **Sohini Roychowdhury**
8:45am - 9:05am	Distributed Rank-1 Dictionary Learning: Towards Fast and Scalable Solutions for fMRI Big Data Analytics	Milad Makkie, Xiang Li, Binbin Lin, Jieping Ye, Tianming Liu, and Shannon Quinn,
9:05am - 9:25am	Simple and Effective Pre-processing for Automated Melanoma Discrimination Based on Cytological Findings	Takuya Yoshida, M.Emre Celebi, Gerald Schaefer, and Hitoshi Iyatomi
9:25am – 9:45am	Mortality Prediction of ICU Patients using Lab Test Data by Feature Vector Compaction & Classification	Mohammad Masud and Abdel Rahman Al Harahsheh
9:45am – 10:05am	Iterative Unified Clustering in Big Data	Vasundhara Misal, Vandana Janeja, Sai Pallaprolu, Yelena Yesha, and Raghu Chintalapati
10:05am - 10:20am	**Coffee Break**	

10:20am - 10:40am	Wearable Sensor based Human Posture Recognition	Jianwu Wang, Zhichuan Huang, Wenbin Zhang, Ankita Patil, Ketan Patil, Ting Zhu, Eric Shiroma, Mitchell Schepps, and Tamara Harris
10:40am - 11:00am	A Framework to Predict Outcome for Cancer Patients Using Data from a Nursing EHR	Muhammad Lodhi, Rashid Ansari, Yingwei Yao, Gail Keenan, Diana Wilkie, and Ashfaq Khokhar
11:00am - 11:20am	Big Data Approach in Healthcare Used for Intelligent Design	Weider Yu, Jaspal Gill, Maulin Dalal, Piyush Jha, and Sajan Shah
11:20am – 12:00pm	**Panel Session: Recent Advancements and Trends in Big Data Analytics for Healthcare** **Panel Chairs: Donghui Wu and Xin Deng**	

W21: The 3rd International Workshop on Pattern Mining and Application of Big Data (BigPMA 2016)		
Time	**Title**	**Presenter/Author**
8:40am - 10:00am (20mins for each)	**Session 1** Chair: Wen-Yuan Zhu	
	A Markov Chain Collaborative Filtering Model for Course Enrollment Recommendations	Elham Khorasani, Zhao Zhenge, and John Champaign
	Using Semantic-based Approach to Manage Perspectives of Process Mining: Application on Improving Learning Process Domain Data	Okoye Kingsley, Abdel-Rahman Tawil, Usman Naeem, Syed Islam, and Elyes Lamine
	Universal Data Discovery Using Atypicality	Anders Host-Madsen, Elyas Sabeti, Chad Walton, and Su Jun Lim
	Topic Modeling for Management Sciences: A Network-based Approach	Max Menenberg, Surya Pathak, Hari Udyapuram, Srinagesh Gavirneni, and Sohini Roychowdhury
10 :00am - 10:20pm	**Coffee Break**	
10:20am - 12:00pm (20mins for each)	Leveraging Cloud Data to Mitigate User Experience from 'Breaking Bad'	Nicholas James, Arun Kejariwal, and David Matteson
	The Technical Hashtag in Twitter Data: a Hadoop Experience	Izabela Moise
	Label Propagation in Big Data to Detect Remote Access Trojans	Sai Pallaprolu, Josephine Namayanja, Vandana Janeja, and Sai Adithya
	Exploring the utilization of places through a scalable "Activities in Places" analysis mechanism	Linlin You, and Bige Tuncer
	An Efficient Parallel Topic-Sensitive Expert Finding Algorithm Using Spark	Yao-Ming Yang, Chang-Dong Wang, and Jian-Huang Lai
12 :00pm - 1:40pm	**Lunch Break**	
1:40pm - 3:20 pm (20mins for each)	**Session 2** Chair: Yi-Cheng Chen	
	A Novel Big-Data Processing Framwork for Healthcare Applications Big-Data-Healthcare-in-a-Box	Fuad Rahman, Marvin Slepian, and Ari Mitra
	Probabilistic Parallelisation of Blocking Non-matched Records for Big Data	Chenxiao Dou, Daniel Sun, Yi-Cheng Chen, Guoqiang Li, and Jianquan Liu
	Interactive Personalized Interesting Pattern Discovery	Mansurul Bhuiyan, and Mohammad Hasan
	Android Malware Detection with Weak Ground Truth Data	Jordan DeLoach, Doina Caragea, and Xinming Ou
	Predicting Traffic of Online Advertising in Real-time Bidding Systems from Perspective of Demand-Side Platforms	Hsu-Chao Lai, Wen-Yueh Shih, Jiun-Long Huang, and Yi-Cheng Chen
3 :20pm - 3:50pm	**Coffee Break**	

W22: Advances in High Dimensional Big Data 2nd Workshop		
Time	**Title**	**Presenter/Author**
1:30pm	Opening	Chair: Sotiris Tasoulis
1:35pm	**Keynote Speech:** Graphical Modeling and the Bethe Approximation	Tony Jebara, Columbia University, USA.
	Paper presentations (1 hour and 15 minutes)	
2:15 pm (25mins for each)	TSmap3D: Browser Visualization of High Dimensional Time Series Data	Supun Kamburugamuve, Pulasthi Wickramasinghe, Saliya Ekanayake, Chathuri Wimalasena, Milinda Pathirage, and Geoffrey Fox
	On the theory and practice of high-dimensional data indexing with iDistance	Ather Sharif, Sarah Cooney, Drew Vitek, and Shengqi Gong Michael Schuh and Rafal Angryk
	Minimum Density Hyperplanes in the Feature Space	Katie Yates and Nicos Pavlidis
3:30 pm	**Coffee Break**	
3:50 pm	**Keynote Speech:** Applied Data Science: Living with the Curse of Dimensionality	Maxime Fournes, Seldon Technologies , UK.
	Paper presentations (1 hour and 15 minutes)	
4:35 pm (25mins for each)	Influence Sketching: Finding Influential Samples In Large-Scale Regressions	Dippy Aggarwal Michael Wojnowicz, Ben Cruz, Xuan Zhao, Brian Wallace, Matt Wolff, Jay Luan, and Caleb Crable
	A Novel Framework for Mitigating Insider Attacks in Big Data Systems Robust K-Subspaces Recovery with Combinatorial Initialization	Santosh Aditham and Nagarajan Ranganathan Jun He, Yue Zhang, Jiye Wang, Nan Zeng, and Hanyong Hao
	Structure Preserving Dimension Reduction with 2D Images as Predictors	Bo Zhang and Liwei Wang
5:50 pm	Closing	Chair: Sotiris Tasoulis

W25: 3rd International Workshop on Privacy and Security of Big Data (PSBD 2016)		
Time	**Title**	**Presenter/Author**
8:00am – 8.25am	**Session PSBD16_1: Opening** **Chair: Alfredo Cuzzocrea**	
8:25am – 9.25am	**Session PSBD16_2: Invited Talk – Elisa Bertino, "Data Privacy for IoT Systems Concepts, Approaches, and Research Directions"** **Chair: Alfredo Cuzzocrea**	
9:25am – 10.45am	**Session PSBD16_3: Secure Methods and Malware Detection Algorithms for Big Data** **Chair: TBA**	
9:25am – 9.45am	Memory Access Pattern based Insider Threat Detection in Big Data Systems	Santosh Aditham, Nagarajan Ranganathan, Srinivas Katkoori
9:45am – 10.05am	Security and Privacy for Big Data: A Systematic Literature Review	Boel Nelson, Tomas Olovsson
10:05am – 10.25am	Reverse Engineering Smart card Malware using Side Channel Analysis with Machine learning Techniques	Hippolyte Djonon Tsague, Twala Bheki
10:25am – 10.45am	Feature Selection and Improving Classification Performance for Malware Detection	Carlos Cepeda, Pablo Ordonez, Dan T-Chia
10:45am - 11:05am	**Coffee Break**	
11:05am – 12.45am	**Session PSBD16_4: Security Frameworks for Supporting Big Data Privacy** **Chair: TBA**	
11:05am – 11.25am	Phishing Through Social Bots on Twitter	Mohammad Shafahi, Leon Kempers, Hamideh Afsarmanesh
11:25am – 11.45am	Phishing Detection Based on Newly Registered Domains	Xueni Li, Guanggang Geng, Zhiwei Yan, Yong Chen,

		Xiaodong Lee
11:45am – 12.05am	Automated Big Security Text Pruning and Classification	Khudran Alzhrani, Ethan Rudd, Edward Chow, Terrance Boult
12:05am – 12.25am	Concise Essence-Preserving Big Data Representation	Philip Derbeko, Shlomi Dolev, Ehud Gudes, Jeffrey Ullman
12:02am – 12.45am	An Entropy-based Analytic Model for the Privacy-Preserving in Open Data	Soo-Hyung Kim, Changwook Jung, Yoon-Joon Lee
12:45am - 2:00pm	**Lunch**	
2:00pm – 4.05pm	**Session PSBD16_5: Privacy-Preserving Big Data Management** **Chair: TBA**	
2:00pm – 2.20pm	Private Databases on the Cloud: Models, Issues and Research Perspectives	Alfredo Cuzzocrea, Carlo Mastroianni, Giorgio Mario Grasso
2:20pm – 2.40pm	Trusted Cloud SQL DBS with On-the-fly AES Decryption/Encryption	Sushil Jajodia, Witold Litwin, Thomas Schwarz
2:40pm – 3.00pm	S3C: An Architecture for Space-Efficient Semantic Search over Encrypted Data in the Cloud	Jason Woodworth, Mohsen Amini Salehi, Vijay Raghavan
3:00pm – 3.20pm	Big Data Analytics as-a-Service: Issues and challenges	Claudio Ardagna, Paolo Ceravolo, Ernesto Damiani
3:20pm – 4.05pm	**Session PSBD16_6: Panel: "Privacy Issues of Big Data Management: What's Next?"** **Chair: Alfredo Cuzzocrea** **Panel Members: TBA**	
4:05pm - 4:25pm	**Coffee Break**	

W26: IEEE Workshop On Big Data Analytics In Manufacturing And Supply Chains

Time	Title	Presenter/Author
1:30 Pm - 1:40 Pm	Opening Remarks	
1:40 Pm - 2:00 Pm	S26204: "Forecast Upc-Level Fmcg Demand: Part Iii: Grouped Reconciliation"	Dr Dazhi Yang
2:00 Pm - 2:20 Pm	S26214: "Prediction Of Regional Goods Demand Incorporating The Effect Of Weather"	Takuya Watanabe
2:20 Pm – 2:40 Pm	S26209: "The Bayesian Estimators Of Polytomous Item Response Theory Models With Approximated Conditional Likelihood And Their Mathematical Optimarities"	Kazumasa Mori
2:40pm - 3:00pm	S26215: "Deep Learning In The Automotive Industry: Applications And Tools"	Andre Luckow
3:00pm- 3:20pm	S26207: "Weighted Clustering Of Spatial Pattern For Optimal Logistics Hub Deployment"	Dr Rong Wen
3:30 Pm - 3:50 Pm	**Coffee Break**	
3:50 Pm - 4:10 Pm	S26210: "Vessel Movement Analysis And Pattern Discovery Using Density-Based Clustering Approach"	Dr Rong Wen
4:10 Pm - 4:30 Pm	S26212: "Adaptive Resilient Strategies For Supply Chain Networks"	Wen Jun Tan
4:30 Pm – 4:50 Pm	S26208: "A Systems Approach To Big Data Technology Applied To Supply Chain"	Tomohiro Fukui
4:50 Pm - 5:30 Pm	**Poster Presentation And Networking**	
	S26202: "Spatial Data Dimension Reduction Using Quadtree: A Case Study On Satellite-Derived Solar Radiation"	*Dr Dazhi Yang*
	S26203: "Analysis For Supply Hub In Industrial Cluster: Classic Vs. New Perspective"	*Vahid Kayvanfar*
	S26205: "A Dea Approach For Supplier Selection With Ahp And Risk Consideration"	*Jasmine Jiamin Lim*
	S26206: "Data Blending In Manufacturing And Supply Chains"	*Boon Yong Ong*
	S26211: "Optimizing Performance Of Sentiment Analysis Through Design Of Experiments"	Gary Goh

W27: Workshop on Textual Customer Feedback Mining and Transfer Learning		
Time	**Title**	**Presenter/Author**
8:00am – 8:30am	Opening Remarks	Xin Deng
8:30am – 9:00am	Keynote Talk I "Improving the Customer Experience by Listening to Data"	Ross Smith
9:00am – 9:30am	Keynote Talk II "Transfer Learning and its Applications on Social Media"	Ming Shao
9:30am – 9:50am	Big Social Data Analytics of Changes in Consumer Behaviour and Opinion of a TV Broadcaster	Anna Hennig, Anne-Sofie Åmodt, Henrik Hernes, Helene Mejer Nygårdsmoen, Peter Arenfeldt Larsen, Raghava Rao Mukkamala, Benjamin Flesch, Abid Hussain, and Ravi Vatrapu
9:50am – 10:15am	**Coffee Break**	
10:15am – 10:30am	Word Embeddings for Arabic Sentiment Analysis	A. Aziz Altowayan and Lixin Tao
10:30am – 10.45am	Giving Voice to Office Customers	Michael Bentley and Soumya Batra
10:45am – 11:00am	Content-based Recommendation for Podcast Audio-items using Natural Language Processing Techniques	Zhou Xing
11:00am –11:15am	Totally Automated Keyword Extraction	Tayfun Pay
11:15am – 11:30am	Unlock Big Data Emotions: Weighted Word Embeddings for Sentiment Classification	Xiangfeng Dai and Bob Prout
11:30am – 11:50am	TV Ratings vs. Social Media Engagement: Big Social Data Analytics of the Scandinavian TV Talk Show Skavlan	Henrikke Hovda Larsen, Johanna Margareta Forsberg, Sigrid Viken Hemstad, Raghava Rao Mukkamala, Abid Hussain, and Ravi Vatrapu

W28: Big NLP 2016		
Time	**Title**	**Authors**
1:30pm – 1:30pm	**Welcome and Introduction (Programme Chair: Paul Rayson)**	
	Session 1: Social/topic and large-scale processing	
1:30pm - 1:50pm	Domain-specific user preference prediction based on multiple user activities	Yunfei Long, Qin Lu, Yue Xiao, MingLei Li, and Chu-Ren Huang
1:50pm - 2:10pm	Large-scale text processing pipeline with Apache Spark	Alexey Svyatkovskiy, Kosuke Imai, Mary Kroeger, and Yuki Shiraito
2:10pm - 2:30pm	lexiDB: A Scalable Corpus Database Management System	Matthew Coole, Paul Rayson, and John Mariani
	Session 2: Annotation	
2:30pm – 2:50pm	Scaling Character-Based Morphological Tagging to Fourteen Languages	Georg Heigold, Günter Neumann, and Josef van Genabith
2:50pm – 3:10pm	A Grapheme-level Approach for Constructing a Korean Morphological Analyzer without Linguistic Knowledge	Jihun Choi, Jonghem Youn, and Sang-goo Lee
3:10pm – 3:30pm	Lightweight System for NE-tagged News Headlines corpus creation	Avinash Kumar, Dhaval Patel, and Nikita Jain
3:30m - 3:50pm	**Afternoon Coffee Break**	
	Session 3: Classification	
3:50pm - 4:10pm	Document Classification through Image-Based Character Embedding and Wildcard Training	Daiki Shimada, Ryunosuke Kotani, and Hitoshi Iyatomi
4:10pm - 4:30pm	Automatic Classification of Securities using Hierarchical Clustering of the 10-Ks	Hoseong Yang, Hye Jin Lee, Eugene Cho, and Sungzoon Cho
4:30pm – 4:50pm	Large-Scale Taxonomy Categorization for Noisy Product Listings	Pradipto Das, Yandi Xia, Aaron Levine, Giuseppe Di Fabbrizio, and Ankur Datta
4:50pm - 5:10pm	Efficient Natural Language Pre-processing for analysing large data sets	Billal Belainine, Alexsandro Fonseca, and Fatiha Sadat
5:10pm – 6:00pm	**System demonstrations**	

Posters

Paper ID	Accept Poster
P201	Priyanka Kale and Shilpa Balan, *Big Data Application in Job Trend Analysis* **Author Email(s):** sbalan@calstatela.edu, pkale@calstatela.edu **Contact Person:** Shilpa Balan <sbalan@calstatela.edu>
P202	raja boddu, *An Integrated Assessment Approach to different Collaborative Filtering Algorithms* **Author Email(s):** iamsarathphd@gmail.com **Contact Person:** raja boddu <iamsarathphd@gmail.com>
P203	Ling He and Jiebo Luo, *What Makes a Pro Eating Disorder Hashtag: Using Hashtags to Identify Pro Eating Disorder Tumblr Posts and Twitter Users* **Author Email(s):** lhe4@u.rochester.edu, jiebo.luo@gmail.com **Contact Person:** Ling He <lhe4@u.rochester.edu>
P204	Vivian Lai, Kyong Jin Shim, Richard Jayadi Oentaryo, Philips Kokoh Prasetyo, Casey Vu, Ee-Peng Lim, and David Lo, *CareerMapper: An Automated Resume Evaluation Tool* **Author Email(s):** kjshim@smu.edu.sg, vivian.lai@colorado.edu, roentaryo@smu.edu.sg, pprasetyo@smu.edu.sg, vuanhthu888@gmail.com, eplim@smu.edu.sg, davidlo@smu.edu.sg **Contact Person:** Kyong Jin Shim <kjshim@smu.edu.sg>
P205	Antonette Shibani, Elizabeth Koh, Vivian Lai, and Kyong Jin Shim, *Analysis of Teamwork Dialogue: A Data Mining Approach* **Author Email(s):** kjshim@smu.edu.sg, antonette.x@nie.edu.sg, elizabeth.koh@nie.edu.sg, vivian.lai@colorado.edu **Contact Person:** Kyong Jin Shim <kjshim@smu.edu.sg>
P210	Ranjeet Devarakonda, Yaxing Wei, and Michele Thornton, *Accessing and Distributing Large Volumes of NetCDF Data* **Author Email(s):** DEVARAKONDAR@ORNL.GOV, weiy@ornl.gov, thorntonmm@ornl.gov **Contact Person:** Ranjeet Devarakonda <DEVARAKONDAR@ORNL.GOV>
P211	ANKUR PADIA, Konstantinos Kalpakis, and Tim Finin, *Inferring Relations in Knowledge Graphs with Tensor Decompositions* **Author Email(s):** ankurpadia@umbc.edu, kalpakis@umbc.edu, finin@umbc.edu **Contact Person:** ANKUR PADIA <ankurpadia@umbc.edu>
P212	Srabasti Dutta, *Is There A Correlation Between Weather and Weather Related Tweets* **Author Email(s):** srabastidutta@gmail.com **Contact Person:** Srabasti Dutta <srabastidutta@gmail.com>
P213	Xingang Wang, *An Approach for Extracting Big Micro-Scale Severe Weather Region Trajectories Automatically from Meteorological Radar Data* **Author Email(s):** wangxingang2009@hotmail.com **Contact Person:** Xingang Wang <wangxingang2009@hotmail.com>
P214	Ranjeet Devarakonda, Kyle Dumas, Sherman Beus, Everett Rush, Bhargavi Krishna, Robert Records, and Giri Prakash, *Next-Gen Tools for Big Scientific Data: ARM Data Center Example* **Author Email(s):** DEVARAKONDAR@ORNL.GOV, dumaskk@ornl.gov, Sherman.Beus@pnnl.gov, rusheniii@ornl.gov, krishnab@ornl.gov, recordsrj@ornl.gov, palanisamyg@ornl.gov **Contact Person:** Ranjeet Devarakonda <DEVARAKONDAR@ORNL.GOV>

P215	Hiroki Imabayashi, Yu Ishimaki, Akira Umayabara, and Hayato Yamana, *Fast and Space-Efficient Secure Frequent Pattern Mining by FHE* **Author Email(s):** imabayashi@yama.info.waseda.ac.jp, yuishi@yama.info.waseda.ac.jp, uma@yama.info.waseda.ac.jp, yamana@yama.info.waseda.ac.jp **Contact Person:** Hiroki Imabayashi <imabayashi@yama.info.waseda.ac.jp>
P216	Ricky Laishram, Katchaguy Areekijseree, and Sucheta Soundarajan, *Predicted Max Degree Sampling : Sampling in Directed Networks to Maximize Node Coverage through Crawling* **Author Email(s):** rlaishra@syr.edu, kareekij@syr.edu, susounda@syr.edu **Contact Person:** Ricky Laishram <rlaishra@syr.edu>
P217	Katchaguy Areekijseree, Ricky Laishram, and Sucheta Soundarajan, *Max-Node Sampling: an Expansion-Densification Algorithm for Data Collection* **Author Email(s):** kareekij@syr.edu, rlaishra@syr.edu, susounda@syr.edu **Contact Person:** Katchaguy Areekijseree <kareekij@syr.edu>
P218	Shaunak Bopardikar and George Ekladious, *Sequential Randomized Matrix Factorization for Gaussian Processes* **Author Email(s):** ekladigs@utrc.utc.com, bopardsd@utrc.utc.com **Contact Person:** George Ekladious <ekladigs@utrc.utc.com>
P220	David Kimmey and Jin Soung Yoo, *Nowcasting with Social Media Data* **Author Email(s):** yooj@ipfw.edu, kimmd01@ipfw.edu **Contact Person:** Jin Soung Yoo <yooj@ipfw.edu>
P221	Ayae Ichinose, Atsuko Takefusa, Hidemoto Nakada, and Masato Oguchi, *Evaluation of Distributed Processing of Caffe Framework Using Poor Performance Device* **Author Email(s):** g1220504@is.ocha.ac.jp, takefusa@nii.ac.jp, hide-nakada@aist.go.jp, oguchi@is.ocha.ac.jp **Contact Person:** Ayae Ichinose <g1220504@is.ocha.ac.jp>
P222	Quanzhi Li, Sameena Shah, Mohammad Ghassemi, Rui Fang, Armineh Nourbakhsh, and Xiaomo Liu, *Using Paraphrases to Improve Tweet Classification: Comparing WordNet and Word Embeddings Approaches* **Author Email(s):** quanzhi.li@thomsonreuters.com, sameena.shah@thomsonreuters.com, ghssemi@mit.edu, rui.fang@thomsonreuters.com, armineh.nourbakhsh@thomsonreuters.com, xiaomo.liu@thomsonreuters.com **Contact Person:** Quanzhi Li <quanzhi.li@thomsonreuters.com>
P224	Vy Bui, Lin-Ching Chang, Dunling Li, Li-Yueh Hsu, and Marcus Chen, *Comparison of Lossless Video and Image Compression Codecs for Medical Computed Tomography Datasets* **Author Email(s):** 01bui@cua.edu, changl@cua.edu, dunling.li@gmail.com, lyhsu@nhlbi.nih.gov, chenmy@nhlbi.nih.gov **Contact Person:** Vy Bui <01bui@cua.edu>
P225	Benito Perez, Xiaomeng Liang, Negin Askarzadeh, Mengran Wang, and Yiwei Ma, *Towards a More Meterless Parking System: Understanding Meter Payment Behavior and Trends in Washington, DC* **Author Email(s):** benito.perez@dc.gov, amy.liang@dc.gov, negin.askarzadeh@dc.gov, mengran.wang@dc.gov, yiwei.ma@dc.gov **Contact Person:** Benito Perez <benito.perez@dc.gov>
P226	Jonathan Rogers, Soumya Dey, Richard Retting, Rahul Jain, Xiaomeng Liang, and Negin Askarzadeh, *Using Automated Enforcement Data to Achieve Vision Zero Goals: A Case Study* **Author Email(s):** soumya.dey@dc.gov, jonathan.rogers@dc.gov, rretting@samschwartz.com, rahul.Jain@dc.gov, amy.liang@dc.gov, negin.askarzadeh@dc.gov **Contact Person:** Soumya Dey <soumya.dey@dc.gov>

P227	Guangxia Xu, Jingteng Zhao, and Deling Huang, *An Improved Social Spammer Detection Based on Tri-training* **Author Email(s):** xugx@cqupt.edu.cn, 403019531@qq.com, huangdl@cqupt.edu.cn **Contact Person:** Guangxia Xu <xugx@cqupt.edu.cn>
P228	Giri Prakash, Jitendra Kumar, Everett Rush, Robert Records, Anthony Clodfelter, and Jimmy Voyles, *HPC Infrastructure to Support the Next-Generation ARM Facility Data Operations* **Author Email(s):** palanisamyg@ornl.gov, kumarj@ornl.gov, rusheniii@ornl.gov, recordsrj@ornl.gov, clodfelteraj@ornl.gov, jimmy.voyles@pnnl.gov **Contact Person:** Giri Prakash <palanisamyg@ornl.gov>
P230	Peter Bajcsy, Soweon Yoon, Piotr M. Szczypinski, Mylene Simon, Mary Brady, Ram Sriram, Stephen J. Florczyk, Nathan Hotaling, Nicholas Schaub, and Carl G. Simon, *Modeling, Validation and Verification of Cell-Scaffold Contact Measurements over Terabyte-sized 3D Image Collection* **Author Email(s):** peter.bajcsy@nist.gov, soweon.yoon@nist.gov, piotr.szczypinski@p.lodz.pl, mylene.simon@nist.gov, mary.brady@nist.gov, ram.sriram@nist.gov, Stephen.Florczyk@ucf.edu, nathan.hotaling@nist.gov, nicholas.schaub@nist.gov, carl.simon@nist.gov **Contact Person:** Peter Bajcsy <peter.bajcsy@nist.gov>
P231	Xiaomeng Liang, Lin-Ching Chang, and Arash Massoudieh, *A Framework for Large-scale Bacterial Motility Behavior Analysis* **Author Email(s):** 74liang@cua.edu, changl@cua.edu, massoudieh@cua.edu **Contact Person:** Xiaomeng Liang <74liang@cua.edu>
P232	Akira Ishii, Masanori Ajito, and Yasuko Kawahata, *Analysis of PokemonGO using sociophysics approach* **Author Email(s):** ishii@damp.tottori-u.ac.jp, purplemukadesan@gmail.com **Contact Person:** Akira Ishii <ishii@damp.tottori-u.ac.jp>
P233	Kenneth David Strang and Zhaohao Sun, *Meta-Analysis of Big Data Security and Privacy: Scholarly Literature Gaps* **Author Email(s):** kenneth.strang@gmail.com, zhaohao.sun@gmail.com **Contact Person:** Kenneth David Strang <kenneth.strang@gmail.com>
P234	Jeffrey Jenkins, Lin-Ching Chang, Elizabeth Hutchinson, M. Okan Irfanoglu, and Carlo Pierpaoli, *Harmonization of Methods to Facilitate Reproducibility in Medical Data Processing: Applications to Diffusion Tensor Magnetic Resonance Imaging* **Author Email(s):** 54jenkins@cua.edu, changl@cua.edu, elizabeth.hutchinson@nih.gov, irfanogl@gmail.com, cp1a@nih.gov **Contact Person:** Jeffrey Jenkins <54jenkins@cua.edu>
P235	Yu Ishimaki, Hiroki Imabayashi, Kana Shimizu, and Hayato Yamana, *Privacy-Preserving String Search for Genome Sequences with FHE bootstrapping optimization* **Author Email(s):** yuishi@yama.info.waseda.ac.jp, imabayashi@yama.info.waseda.ac.jp, shimizu.kana@waseda.jp, yamana@yama.info.waseda.ac.jp **Contact Person:** Yu Ishimaki <yuishi@yama.info.waseda.ac.jp>
P237	Adel Assiri, *Real-Time Sentiment Analysis of Saudi Dialect Tweets Using SPARK* **Author Email(s):** aadel_3@hotmail.com **Contact Person:** Adel Assiri <aadel_3@hotmail.com>
P238	Seungwoo Jeon, Jaegi Hong, Bonghee Hong, and Chumsu Kim, *TPR*-tree Performance Improvement for Big Tactical Moving Objects* **Author Email(s):** bhhong@pusan.ac.kr, i2825t@pusan.ac.kr, hjg5921@pusan.ac.kr, chskim@add.re.kr **Contact Person:** Bonghee Hong <bhhong@pusan.ac.kr>

P239	Austin Harris, Hanna True, Zhen Hu, Jin Cho, Nancy Fell, and Mina Sartipi, *Fall Recognition using Wearable Technologies and Machine Learning Algorithms* **Author Email(s):** Mina-Sartipi@utc.edu, sbn486@mocs.utc.edu, mdq598@mocs.utc.edu, zhen-hu@utc.edu, cwp636@mocs.utc.edu, Nancy-Fell@utc.edu **Contact Person:** Mina Sartipi <Mina-Sartipi@utc.edu>
P240	Jiwan Lee, Jaegi Hong, Bonghee Hong, and Jinsu Ahn, *A Generator of Test Data Set for Tactical Moving Objects Based on Velocity* **Author Email(s):** bhhong@pusan.ac.kr, wldhks85@pusan.ac.kr, hjg5921@pusan.ac.kr **Contact Person:** Bonghee Hong <bhhong@pusan.ac.kr>
P241	Xiaoxia Jia, Peng Cheng, and Jiming Chen, *A Data Analysis and Visualization System for Large-Scale e-Bike Data* **Author Email(s):** saodiseng@gmail.com, xiaoxiajia@zju.edu.cn, cjm@zju.edu.cn **Contact Person:** Peng Cheng <saodiseng@gmail.com>
P243	Sunghwan Cho, Sunghak Hong, and Changsoo Lee, *ORANGE: Spatial Big Data Analysis and Visualization Platform* **Author Email(s):** mission21@gmail.com **Contact Person:** Sunghwan Cho <mission21@gmail.com>

Hyatt Regency Washington on Capitol Hill Floor Plan

HYATT REGENCY WASHINGTON
ON CAPITOL HILL
400 New Jersey Avenue, N.W.
Washington, D.C. 20001
USA

T +1 202 737 1234
Sales Fax +1 202 719 8419
Guest Fax +1 202 737 5773
qualitywasrw@hyatt.com
washingtonregency.hyatt.com

FLOOR PLAN
All Floors

Conference WiFi Instruction

2016 Big Data Conference

Wifi Access

Connect to Wireless Network: "Hyatt-Meeting"

SSID, Open Explorer and you'll be on PSAV page;

Access Code (left side of screen): IEEEBD2016 (case sensitive)

Select: "I agree to the terms and conditions"

IEEE computer society

IEEE BIGDATA 2017

December, 2017, Philadelphia, PA, USA

The IEEE Big Data conference series is a leading forum for disseminating the latest advances in big data research , development and application. We solicit high-quality original research papers (including significant work-in-progress) in any aspect of Big Data with emphasis on 5Vs (Volume, Velocity, Variety, Value and Veracity): big data science and foundations, big data infrastructure, big data management, big data searching and mining, big data privacy/security, and big data applications. Relevant topics include but are not limited to:

1. Big Data Science and Foundations

a. Novel Theoretical Models for Big Data
b. New Computational Models for Big Data
c. Data and Information Quality for Big Data
d. New Data Standards

2. Big Data Infrastructure

a. Cloud/Grid/Stream Computing for Big Data
b. High Performance/Parallel Computing Platforms for Big Data
c. Autonomic Computing and Cyber-infrastructure, System Architectures, Design and Deployment
d. Energy-efficient Computing for Big Data
e. Programming Models and Environments for Cluster, Cloud, and Grid Computing to Support Big Data
f. Software Techniques and Architectures in Cloud/Grid/Stream Computing
g. Big Data Open Platforms
h. New Programming Models for Big Data beyond Hadoop/MapReduce, STORM
i. Software Systems to Support Big Data Computing

3. Big Data Management

a. Advanced database and Web Applications
b. Novel Data Model and Databases for Emerging Hardware
c. Data Preservation
d. Data Provenance
e. Interfaces to Database Systems and Analytics Software Systems
f. Data Protection, Integrity and Privacy Standards and Policies
g. Information Integration and Heterogeneous and Multi-structured Data Integration
h. Data management for Mobile and Pervasive Computing
i. Data Management in the Social Web
j. Crowdsourcing
k. Spatiotemporal and Stream Data Management
l. Scientific Data Management
m. Workflow Optimization
n. Database Management Challenges: Architecture, Storage, User Interfaces

4. Big Data Search and Mining

a. Social Web Search and Mining
b. Web Search
c. Algorithms and Systems for Big Data Search
d. Distributed, and Peer-to-peer Search
e. Big Data Search Architectures, Scalability and Efficiency
f. Data Acquisition, Integration, Cleaning, and Best Practice
g. Visualization Analytics for Big Data
h. Computational Modeling and Data Integration
i. Large-scale Recommendation Systems and Social Media Systems
j. Cloud/Grid/Stream Data Mining- Big Velocity Data
k. Link and Graph Mining
l. Semantic-based Data Mining and Data Pre-processing
m. Mobility and Big Data
n. Multimedia and Multi-structured Data- Big Variety Data

5. Big Data Security & Privacy

a. Intrusion Detection for Gigabit Networks
b. Anomaly and APT Detection in Very Large Scale Systems
c. High Performance Cryptography
d. Visualizing Large Scale Security Data
e. Threat Detection using Big Data Analytics
f. Privacy Threats of Big Data
g. Privacy Preserving Big Data Collection/Analytics
h. HCI Challenges for Big Data Security & Privacy
i. User Studies for any of the above
j. Sociological Aspects of Big Data Privacy

6. Big Data Applications

a. Complex Big Data Applications in Science, Engineering, Medicine, Healthcare,Finance, Business, Law, Education, Transportation, Retailing, Telecommunication
b. Big Data Analytics in Small Business Enterprises (SMEs),
c. Big Data Analytics in Government, Public Sector and Society in General
d. Real-life Case Studies of Value Creation through Big Data Analytics
e. Big Data as a Service
f. Big Data Industry Standards
g. Experiences with Big Data Project Deployments

INDUSTRIAL and GOVERNMENT Track

The Industrial and government Track solicits papers describing implementations of Big Data solutions relevant to industrial settings. The focus of industry track is on papers that address the practical, applied, or pragmatic or new research challenge issues related to theuse of Big Data in industry. We accept full papers (up to 10 pages) and extended abstracts (2-4 pages.