

**IEEE**

**BIG DATA 2019**

Los Angeles, CA, USA • 9–12 December

*Sponsored by*



**very**



# 2019 IEEE International Conference on Big Data

<b>Organization Committee.....</b>	<b>2</b>
<b>Program Committee .....</b>	<b>4</b>
<b>IEEE Big Data 2019 Program Schedule.....</b>	<b>14</b>
<b>Keynote Lectures .....</b>	<b>22</b>
<b>Conference Paper Presentations.....</b>	<b>25</b>
<b>Industry and Government Paper Presentations .....</b>	<b>34</b>
<b>Tutorials .....</b>	<b>36</b>
<b>Workshops.....</b>	<b>40</b>
<b>Special Symposiums .....</b>	<b>63</b>
<b>Special Sessions .....</b>	<b>64</b>
<b>BigData Cup Challenges .....</b>	<b>71</b>
<b>Panel.....</b>	<b>73</b>
<b>Posters.....</b>	<b>74</b>
<b>Conference Wifi Access .....</b>	<b>78</b>
<b>Westin Hotel Floor Plan.....</b>	<b>79</b>
<b>IEEE Big Data 2020.....</b>	<b>81</b>

# Organization Committee

## Conference Co-Chairs

Dr. Roger Barga: Amazon.com, USA

Prof Carlo Zaniolo: UCLA, USA

## Program Co-Chairs

Dr. Chaitanya Baru: San Diego Supercomputer Center/Univ. of California San Diego, USA

Dr. Jun (Luke) Huan: Styling AI Inc., China

Prof. Latifur Khan: University of Texas at Dallas, USA

## Vice Chairs in Big Data Science and Foundations

Prof. Jingrui He: UIUC, USA

Prof. Wenqing Hu: Missouri S&T University, USA

## Vice Chairs in Big Data Infrastructure

Prof. Hanghang Tong: UIUC, USA

Dr. Yinglong Xia: Facebook AI, Menlo Park, CA, USA

## Vice Chairs in Big Data Management

Prof. Christopher Jermaine: Rice University, USA

Prof. Yongluan Zhou: Univ. of Copenhagen, Denmark

## Vice Chairs in Big Data Search and Mining

Prof. Quanquan Gu: UCLA, USA

Prof. Aditya Prakash: Virginia Tech, USA

## Vice Chairs in Big Data Security, Privacy and Trust

Prof. Dongwon Lee: Penn State University, USA

Prof. Julia Stoyanovich: New York University, USA

## Vice Chairs in Hardware/OS Accelerating for Big Data

Prof. Sang-Woo Jun: UC Irvine, USA

Prof. Harry Xu: UCLA, USA

## Vice Chairs in Big Data Applications

Prof. Xia Ning: Ohio State University, USA

Prof. Tim Weninger: Univ. of Notre Dame, USA

## Industry and Government Program Committee Co-Chairs

Dr. Ronay Ak: NVIDIA, USA

Dr. Yuanyuan Tian: IBM Almaden Research Center, USA

## Workshop Co-Chairs

Prof. Kisung Lee: Louisiana State University, USA

Prof. Yanfang Fanny Ye: West Virginia University, USA

## Tutorial Co-Chairs

Prof. Xia (Ben) Hu: Texas A&M Univ, USA

Prof. Chen Li: UC Irvine, USA

Prof. Jeffrey Saltz: Syracuse University, USA

## Big Data Cup Chairs

Prof. Yicheng Tu: South Florida University, USA

## **Poster Co-Chairs**

Prof. Anna Cinzia Squicciarini: Penn State University, USA  
Dr. Haoyi Xiong: Baidu, China  
Prof. Sihong Xie: Lehigh University, USA  
Dr. Lingfei Wu: IBM Research, USA

## **Sponsorship Chairs**

Prof. Xiaohua Tony Hu: Drexel University, USA  
Dr. Salloum Mariam: UC Riverside, USA

## **Local Arrangement Chairs**

Dr. Wei Ye: UC Santa Barbara, USA

## **Registration Chair**

Prof. Yuan An: Drexel University, USA

## **Publicity Chairs**

Prof. Ruoming Jing: Kent State University, USA  
Prof. Amr Magdy: UC Riverside, USA

## **Student Travel Award Chairs**

Prof. Weimao Ke: Drexel University, USA  
Dr. Christine Kirkpatrick: San Diego Supercomputer Center, USA  
Prof. Jianwu Wang: University of Maryland, Baltimore County, USA

## **BigData Steering Committee**

Dr. Amr Awadallah: Cloudera, USA  
Dr. Xueqi Cheng: Chinese Academy of Science, China  
Prof. Yi-ke Guo: Imperial College, UK  
Prof. Xiaohua Tony Hu (Chair) (xh29@drexel.edu): Drexel University, USA  
Prof. Jimmy Lin: University of Waterloo, Canada  
Dr. Raghunath Nambiar: AMD, USA  
Prof. Jian Pei: Simon Fraser University, Canada  
Prof. Vijay Raghavan: University of Louisiana at Lafayette, USA  
Prof. Amit Sheth: University of South Carolina, USA  
Prof. Matthew Smith: Leibniz Universität Hannover, Germany  
Dr. Shusaku Tsumoto: Shimane University, Japan  
Prof. Athanasios Vasilakos: Lulea University of Technology(LTU), Sweden  
Pro. Wei Wang: University of California at Los Angeles, USA  
Prof. Qiang Yang: Hong Kong University of Science and Technology, China

# Program Committee

## Main Conference PC members

Name	Organization	Country/State
Abdelmounaam Rezgui	New Mexico Tech	USA
Abdullah Mueen	University of New Mexico	USA
Abhishek Chandra	University of Minnesota	USA
Abolfazl Asudeh	University of Michigan	USA
Abusayeed Saifullah	Wayne State University	USA
Aibek Musaev	The University of Alabama	USA
Aidong Zhang	University of Virginia	USA
Ajitesh Srivastava	University of Southern California	USA
Aki-Hiro Sato	Kyoto University	Japan
Alex Delis	Univ. of Athens and NYU Abu Dhabi	Greece
Alexander Jung	Aalto University	Finland
Alexander Lazovik	University of Groningen	Netherlands
Alexandros Labrinidis	University of Pittsburgh	USA
Alexandru Costan	IRISA / INSA Rennes	France
Alexey Lastovetsky	University College Dublin	Ireland
Ali R. Butt	Virginia Tech	USA
Alkis Simitsis	HPE Labs	USA
Allan Nengsheng Zhang	Singapore Institute of Manufacturing Technology	Singapore
Amarnath Gupta	San Diego Supercomputing Center	USA
Amelie Chi Zhou	Shenzhen University	China
Amey Kulkarni	FPGA Engineer, DSP and Computer Vision Focus, Velodyne LiDAR, Inc.	USA
Amirreza Masoumzadeh	SUNY at Albany	USA
An Liu	Soochow University	China
Anastasios Gounaris	Aristotle University of Thessaloniki	Greece
Andrea Cali	University of Oxford, UK	UK
Andrea Clematis	IMATI - CNR	Italy
Angela Bonifati	University of Lyon	France
Ankit Agrawal	Northwestern University	USA
Antonio Badia	University of Louisville	USA
Arash Jalal Zadeh Fard	Microfocus - Vertica	USA
Arindam Pal	TCS Research and Innovation	India
Ariyam Das	UCLA	USA
Aylin Caliskan	George Washington University	USA
Azad Naik	Microsoft	USA
Brajendra Nath Panda	University of Arkansas	USA
Balaji Palanisamy	University of Pittsburgh	USA

Baojun Qiu	Chaoda Foodmall Group	China
Barbara Pernici	Politecnico Milano	Italy
Bernd Freisleben	Philipps-Universität Marburg	Germany
Bin Yang	Aalborg University	Denmark
Bishwaranjan Bhattacharjee	IBM Research	USA
Bolong Zheng	Huazhong University of Science and Technology	China
Boris Delibasic	University of Belgrade	Serbia
Boxin Du	ASU	USA
Bradley Malin	Vanderbilt University	USA
Brian D. Davison	Lehigh University	USA
Bruno Schulze	National Lab. for Scientific Computing, Brazil	Brazil
Carlos Varela	Rensselaer Polytechnic Institute	USA
Chao Huang	ND	USA
Chao Lan	University of Wyoming	USA
Chen Haopeng	Shanghai Jiao Tong University	China
Cheng Long	Queen's University Belfast	UK
Chengliang Chai	Tsinghua University	China
Chih-Chieh Yang	IBM Research	USA
Chris Argenta	Applied Research Associates, Inc.	USA
Christan Grant	University of Oklahoma	USA
Christoph Quix	Hochschule Niederrhein, University of Applied Sciences	Germany
Chuanren Liu	Drexel University	USA
Chunqiu Zeng	Google Inc.	USA
Chuxu Zhang	University of Notre Dame	USA
Cinzia Cappiello	Politecnico di Milano	Italy
Clemens Grelck	University of Amsterdam	Netherlands
Da Zhang	University of Miami	USA
Dana Petcu	West University of Timisoara, Romania	Romania
Danilo Ardagna	Politecnico di Milano, Italy	Italy
David Belanger	Stevens Institute of Technology	USA
David Kaeli	Northeastern University	USA
Dawei Zhou	ASU	USA
Deepak Venugopal	University of Memphis	USA
Dimitrios Geogakopoulos	Swinburne U	Australia
Dimitrios Tsoumakos	Department of Informatics, Ionian University	Greece
Dmitriy Fradkin	Siemens	USA
Dmitry Duplyakin	University of Utah	USA
Domenico Talia	University of Calabria	Italy
Dong Chen	Florida International University	USA
Douglas Talbert	Tennessee Technological University	USA
Edgar Meij	University of Amsterdam	Netherlands

Eduard Deagut	Temple University	USA
Edward Curry	Insight Centre for Data Analytics at National University of Ireland, Galway	Ireland
Elisa Bertino	Purdue University	USA
Eric Lo	Chinese University of Hong Kong	Hong Kong
Evaggelia Pitoura	University of Ioannina	Greece
Fabrizio Marozzo	Università della Calabria	Italy
Fang Zhou	Temple University	USA
Felix Gessert	University of Hamburg	Germany
feng lu	Clemson University	USA
Fengguang Song	Indiana University-Purdue University Indianapolis	USA
Frederico Araujo	IBM Research	USA
Gabriele Gianini	Università degli Studi di Milano, Italy and EBTIC at Khalifa University, Abu Dhabi	Italy
George Fletcher	Eindhoven Univ. of Technology	Netherlands
Giovanni Livraga	Università degli Studi di Milano	Italy
Guan-hua Tu	Michigan State University	USA
Guangyou Zhou	Central China Normal University	China
Guangzhong Sun	University of Science and Technology of China	China
Guojie Song	Peking University	China
Guojing Cong	IBM	USA
Guruprasad Nayak	Univ Minnesota	USA
Hai Jin	Huazhong University of Science & Technology, China	China
Hailong Sun	Beihang University	China
Haiqin Yang	Hang Seng Management College	Hong Kong
Haiying Shen	Clemson University	USA
Hamid Karimi	Michigan State University	USA
Hang Liu	Stevens Institute of Technology	USA
Haoyi Xiong	Baidu	China
Heiko Ludwig	IBM Research - Almaden	USA
Hidemoto Nakada	National Institute of Advanced Industrial Science and Technology (AIST)	Japan
Hideyuki Kawashima	Keio University	Japan
Hill Zhu	Florida Atlantic University	USA
Hiroshi Horii	IBM Research - Tokyo	Japan
Hiroshi Inoue	IBM Research - Tokyo	Japan
Hiroshi Mamitsuka	Kyoto University/Aalto University	Japan / Finland
Hoang Bui	Western Illinois University	USA
Hong-Linh Truong	Aalto University	Finland
Hongbo Xu	Chinese Academy of Sciences, China	China
Hongfei Lin	Dalian Univ. of Technology, China	China
Hongfu Liu	Northeastern University	China
Hongwei Liang	Microsoft	Canada



Huasong Shan	JD.com	USA
Huaxiu Yao	PSU	USA
Huiyuan Chen	Case Western Reserve University	USA
Ignacio Arnaldo	PatternEx	USA
Irfan Ahmed	Virginia Commonwealth University	USA
Jaap Kamps	University of Amsterdam	Netherlands
Jack Lange	University of Pittsburgh	USA
Jacob Gao	Pinterest	USA
James Abello	DIMACS/Rutgers, USA	USA
James Joshi	University of Pittsburg, USA	USA
Jan Ramon	INRIA Lille	France
Jane Greenberg	Drexel University	USA
Jay Lofstead	Sandia National Laboratories	USA
Jialong Han	Tencent AI Lab	China
Jian Tang	University of Michigan	USA
Jian Yang	Macquarie University	Australia
Jian-Yun Nie	Univeristy of Montreal	Canada
Jiangfan Zhang	Missouri S&T	USA
Jianguo Wang	Amazon	USA
Jiannong Cao	The Hong Kong Polytechnic University	Hong Kong
Jianpeng Xu	Michigan state university	USA
Jianting Zhang	City College of New York	USA
Jianwen Su	U C Santa Barbara	USA
Jianwu Wang	University of Maryland, Baltimore County	USA
Jimmy Xiangji Huang	York University	Canada
Jinha Kim	Oracle Labs	USA
Joao Eduardo Ferreira	University of Sao Paulo	Brazil
Jorge-Arnulfo Quiane-Ruiz	QCRI	Qatar
Ju Fan	Renmin University of China	China
Juan Zhao	Vanderbilt University	USA
Judy Qiu	Indiana University	USA
Jun Ma	Amazon	USA
Jundong Li	Arizona State University	USA
Junhua Ding	Department of Information Science, University of North Texas	USA
Ka-Chun Wong	City University of Hong Kong	Hong Kong
Kaixiang Lin	Michigan State University	USA
Kazuaki Ishizaki	IBM Research	Japan
Ke Wang	Simon Fraser University	Canada
Keijo Heljanko	University of Helsinki	Finland
Keke Chen	Wright State University	USA
Kenneth Chiu	Binghamton University	USA

Kiyokuni Kawachiya	IBM Research - Tokyo	Japan
Kun Qian	IBM	USA
Kunpeng Zhang	University of Maryland, College Park	USA
Kwan-Liu Ma	University of California, Davis	USA
Lakshmish Ramaswamy	University of Georgia	USA
Larbi Boubchir	University of Paris 8	France
Le Gruenwald	University of Oklahoma	USA
Lecheng Zheng	University of Illinois at Urbana Champaign	USA
Lei Cao	MIT	USA
Li Su	Alibaba	China
Liang Wu	Arizona State University	USA
Lidan Shou	Zhejiang University, China	China
Lifeng Nai	Google	USA
Lijun Chang	The University of Sydney	Australia
Lin-Ching Chang	The Catholic University of America	USA
Liqiang Wang	University of Central Florida	USA
Long Cheng	Dublin City University	Ireland
Magdalini Eirinaki	San Jose State University	USA
Malolan Chetlur	IBM India,	India
Marcos Antonio Vaz Salles	University of Copenhagen - Dep. of Computer Science	Denmark
Mario Bravetti	University of Bologna	Italy
Marios Dikaiakos	University of Cyprus	Cyprus
Martin Berzins	University of Utah	USA
Masatoshi Hanai	Nanyang Technological University	Singapore
Matthias Schunter	Intel	USA
Matthieu Dorier	Argonne National Laboratory	USA
Mauricio Tsugawa	Microsoft	USA
Meihui Zhang	Beijing Institute of Technology	China
Michael Gubanov	Florida State University	USA
Michele Ciavotta	Università Milano Bicocca	Italy
Miki Enoki	IBM Research - Tokyo	Japan
Milos Jovanovic	University of Belgrade	Serbia
Ming Shao	UMass Dartmouth	USA
Ming Zhao	Arizona State University	USA
Ming-Syan Chen	National Taiwan University	Taiwan
Miyuru Dayarathna	WSO2 Inc.	USA
Mohamed Ghalwash	IBM T.J. Watson Research Center	USA
Mohammad Maifi Hasan Khan	University of Connecticut	USA
Mohammad Masud	United Arab Emirates University	USA
Mosharaf Chowdhury	University of Michigan	USA
Muhao Chen	UPenn	USA

Naeemul Hassan	University of Maryland	USA
Nathalie Baracaldo	IBM Almaden Research, USA	USA
Nemanja Djuric	Uber ATG	USA
Nguyen Ho	Aalborg University	Denmark
Nianjun (Joe) Zhou	IBM T.J. Watson Research Center, USA	USA
Nick Duffield	Texas A&M University	USA
Noel De Palma	University Joseph Fourier	France
Pan Li	Case Western Reserve University	USA
Panos Kalnis	KAUST	Saudi Arabia
Paolo Romano	Lisbon University/INESC-ID	Portugal
Paul Grefen	Eindhoven University of Technology Eindhoven University of Technology	Netherlands
Paul Suganthan G C	Google	USA
Peter Baumann	Jacobs University	Germany
Philip Chan	Florida Institute of Technology	USA
Philip Rhodes	University of Mississippi	USA
Philip S. Yu	University of Illinois at Chicago	USA
Pietro Colombo	University of Insubria	Italy
Qi Yu	Rochester Institute of Technology	USA
Qiang Guan	Kent State University	USA
Qingguang Guan	Temple University	USA
Qingsong Wen	Alibaba Group	USA
Ragib Hasan	University of Alabama - Birmingham	USA
Ramin Yahyapour	GWDG/University Goettingen, Germany	Germany
Ran Wolff	Yahoo Research	Israel
Raymond Wong	University of New South Wales	Australia
Ren Chen	Alibaba	USA
Ricardo J. G. B. Campello	University of Newcastle	Australia
Rizos Sakellariou	University of Manchester	UK
Roberto J. Bayardo	Google, USA	USA
Roe E. Ebenstein	Google	USA
Rui Bo	Missouri S&T	USA
Ruoming Jin	Kent State University	USA
Sai Wu	Zhejiang University	China
Sairam Gurajada	IBM	USA
Sajal K. Das	MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY	USA
Salvatore Orlando	Università Ca' Foscari Venezia - Dipartimento di Sc. Ambientali, Informatica e Statistica	Italy
Sanjay Madria	Department of Computer Science, Missouri University Science and Technology	USA
Saso Dzeroski	Jozef Stefan Institute	Slovenia
Sebastien Mosser	Université du Québec à Montréal	Canada
Senzhang Wang	Nanjing University of Aeronautics and Astronautics	China

Shanjiang Tang	Tianjing University	China
Shannon Quinn	University of Georgia	USA
Shaoxu Song	Tsinghua University	China
Sheng Li	University of Georgia	USA
Shinichi Morishita	University of Tokyo	Japan
Shiyong Lu	Wayne State University	USA
Shu-Ching Chen	Florida International University	USA
Sihem Amer-Yahia	CNRS/LIG	France
Sorelle Friedler	Haverford College	Denmark
Suhang Wang	PSU	USA
Sumit Purohit	Pacific Northwest National Laboratory	USA
Suren Byna	Lawrence Berkeley National Lab	USA
Surya Nepal	CSIRO	Australia
Suzanne McIntosh	NYU Courant Institute, and NYU Center for Data Science	USA
Taghi Khoshgoftaar	Florida Atlantic University	USA
Takanori Ueda	IBM Research - Tokyo	Japan
Tamjidul Hoque	University of New Orleans	USA
Taneli Mielikainen	Verizon Media (Yahoo!)	USA
Teruo Higashino	Osaka University	Japan
Ting Wang	<a href="http://x-machine.github.io/">http://x-machine.github.io/</a>	USA
Tiziana Margaria	University of Limerick and Lero - The Irish Software Research Centre	Ireland
Toshio Endo	Tokyo Institute of Technology	Japan
Tyler Derr	Michigan State University	USA
Uwe Roehm	The University of Sydney	Australia
Vahid Taslimitehrani	PhysioSigns Inc.	USA
Vana Kalogeraki	AUEB	Greece
Vanessa Frias-Martinez	University of Maryland	USA
Venkat Srinivasan	Virginia Tech	USA
Victor Junqiu wei	Huawei Noah's Ark Lab	China
Vijay K. Gurbani	Illinois Institute of Technology	USA
Vivian Hu	Ryerson University/University of Calgary	Canada
Vlado Keselj	Dalhousie University	Canada
Walid Aref	Purdue University	USA
Wei Lu	Renmin University	China
Wei Shi	Carleton University	Canada
Weidong Shi	University of Houston	USA
Weijia Xu	University of Texas / Texas Advanced Computing Center	USA
Weiliang Zhao	Macquarie University	Australia
Wenjian Xu	Alibaba	China
Wenjie Zhang	University of New South Wales	Wales
Wonik Choi	Inha University	Korea

Xiang Lian	Kent State University	USA
Xiang Sheng	Syracuse University	USA
Xiangliang Zhang	King Abdullah University of Science and Technology	Saudi Arabia
Xiangyu Zhao	Michigan State University	USA
Xiaofeng Meng	Renmin University of China	China
Xiaoning Ding	New Jersey Inst. of Technology	USA
Xiaoyan Yang	Yitu Singapore	Singapore
Xiaoyong DU	Renmin University of China	China
Xin Yuan	Florida State University	USA
Xiting Wang	Microsoft Research Asia	China
Xuan Zhou	East China Normal University	China
Xueqi Cheng	Chinese Academy of Science	China
Yan Huang	University of North Texas	USA
Yandong Li	university of central florida	USA
Yang Liu	Wilfrid Laurier University	Canada
Yang Zhou	Auburn University	USA
Yanjie Fu	Missouri S&T	USA
Yanmin Zhu	Shanghai Jiao Tong University, China	China
Yao Ma	Michigan State University	USA
Yao Zhou	UIUC	USA
Ye Yuan	Northeastern University	USA
Yiming Ying	The University at Albany, State University of New York	USA
Yinghui Wu	Washington State University	USA
Yingxia Shao	Peking University	China
Yiping Ke	Nanyang Technological University	Singapore
Yong Chen	Texas Tech University	USA
Yu Yang	City University of Hong Kong	Hong Hong
Yuan Yao	Nanjing University	China
Yuchen Bian	The Pennsylvania State University	USA
Yuchen Li	Singapore Management University	Singapore
Yueguo Chen	Renmin University of China, Key Laboratory of Data Engineering and Knowledge Engineering	China
Yusuke Tanimura	National Institute of Advanced Industrial Science and Technology	Japan
Yuzhe Tang	Syracuse University	USA
Zehua Chen	Taiyuan University of Technology	China
Zhanhuai Li	Polytechnical University, Xian	China
Zhanxing Zhu	Peking University	China
Zhengzhang Chen	NEC Laboratories America	USA
Zhi Nie	Samsung Research America	USA
Zhi Yang	Peking University	China
Zhicheng Dou	Renmin University of China	China
Zhilei Xu	Google	USA

Zhiling Lan	Illinois Institute of Technology	USA
Zhining Liu	University of Electronic Science and Technology of China	China
Zhiyuan Chen	University of Maryland Baltimore County	USA
Zhoujun Li	BAUU	China
Zitao Liu	TAL AI Lab	China

## Industry and Government Program PC members

Name	Organization	Country/State
Abdul Quamar	IBM Research - Almaden	USA
Alain crolotte	Teradata Corporation	USA
Anantha Narayanan	University of Maryland, College Park, MD 20742	USA
Antonino Tumeo	Pacific Northwest National Laboratory	USA
Ayca Altay	Rutgers University	USA
Brian Wang	UCLA	USA
Carl Willis-Ford	GDIT	
Carla Freitas Silveira Netto	UFRGS School of Management	Brazil
Chuan Lei	IBM Research - Almaden	USA
Constantinos Makrides	Cray	USA
David José da Silva Aresta Belo	Faculdade de Ciências e Tecnologia – Universidade Nova de Lisboa	Portugal
Erdem Kaya	Sabanci University	Turkey
Frank Soboczinski	King's College London	UK
Gianni Barlacchi	University of Trento	Italy
Haozhen Zhao	Ankura	USA
Harshad Deshmukh	Google	USA
Hasan Huseyin Eruslu	The University of Delaware	USA
James Skinner	NVIDIA	USA
Kai Zeng	Alibaba	China
Koichi Shirahata	Fujitsu	Japan
Konstantinos Karanasos	Microsoft	
Lei Huang	IBM Research - Almaden	USA
Lilong Jiang	Twitter	USA
Lukas Rupprecht	IBM Research - Almaden	USA
Madian Khabza	Apple	USA
Mansurul Bhuiyan	Walmart Labs	USA
Mark-Alexander Henn	NIST	USA
Mehdi Dadfarnia	NIST	USA
Mengwen Liu	Amazon	USA
Michael Sharp	NIST	USA
Mihai Capotă	Intel Labs, Intel Corporation	USA
Min Li	JD.COM	USA

Miyuru Dayarathna	WSO2, Inc.	USA
Mohammed Korayem	Careerbuilder	USA
Mu Qiao	IBM Almaden Research Center	USA
Nancy Grady	SAIC	USA
Nesreen Ahmed	Intel	USA
Olivera Kotevska	ORNL	USA
Pawan Nandakishore	Colaberry Inc	USA
Pranjul Yadav	Criteo AI Lab	USA
Raghav Gupta	Google Research	USA
Ramendra Sahoo	Teradata	USA
Rashi Verma	Tamr Inc	USA
Roberto Rocchetta	National Institute of Aerospace	USA
Rohit Taneja	NVIDIA	USA
Ronay Ak	NVIDIA	USA
Shaden Smith	Intel Labs	USA
Siddha Ganju	NVIDIA	USA
Tatsuhiro Chiba	IBM Research - Tokyo	Japan
Thurston Sexton	National Institute of Standards and Technology	USA
Vasilis Efthymiou	IBM Research - Almaden	USA
Vijay Gadepally	Massachusetts Institute of Technology	USA
Vivian Wen Hui Wong	Stanford University	USA
Wenlei Xie	Facebook	USA
Xia Zhu	Parallel Computing Lab, Intel Corporation	USA
Yida Wang	Amazon	USA
Yingjun Wu	Amazon Web Services	USA
Yuan Ling	Amazon Alexa	USA
Yuanyuan Tian	IBM Research - Almaden	USA
Zhongfang Zhuang	Visa Research	USA

# IEEE Big Data 2019 Program Schedule

Los Angeles, CA, USA

December 9 - December 12, 2019

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 Westin Bonaventure Hotel & Suites located at 404 South Figueroa Street, Los Angeles, CA.

Sunday, December 8, 2019	
3: 00 - 7: 00pm	<b>Registration</b>
Location:	California Foyer

Day 1: Monday, December 9, 2019			
7: 30 - 6: 00 pm	<b>Registration</b>		
Location:	California Foyer		
Time	Sessions/Workshops	Session Chair	Location
Whole Day (between 8 am - 7 pm, please check each individual workshop program schedule)	Workshop: Real-time and Stream Analytics in Big Data & Stream Data Management	Sabri Skhiri, Albert Bifet, and Alessandro Margara	Santa Anita A
	Workshop: Big Spatial Data (BSD)	Abdeltawab Hendawi, Farnoush Banaei-Kashani, Chengyang Zhang, and Siyuan Lu	Santa Anita B
	Workshop: Human-in-the-loop Methods and Human-Machine Collaboration in BigData (HMDData)	Senjuti Basu Roy and Atsuyuki Morishima	Santa Anita C
	Workshop: Big Food and Nutrition Data Management and Analysis (BFNDMA)	Tome Eftimov, Bibek Paudel, and Barbara Koroušić Seljak	San Gabriel A
	Workshop: Big Data for Financial News and Data	Quanzhi Li, Xiaozhong Liu, and Sameena Shah	San Gabriel B
	Special Session: Machine Learning on Big Data (MLDB 2019)	Alfredo Cuzzocrea	Santa Barbara A
Time	Sessions/Workshops	Session Chair	Location
8: 10 - 10: 10	Tutorial 1: Process mining: Leveraging event data to understand and improve organizations	Henrik Leopold, Han van der Aa	San Gabriel C
10: 30 - 12: 30	Tutorial 7: Industrial AI: Machine Learning for Maintenance and Repair	Chetan Gupta, Ahmed Farahat	San Gabriel C
Morning workshop (any time between 8 am - 1 pm, please check each individual workshop schedule)	4 combined Workshops: Advances in High Dimensional (AdHD) Big Data/ Big Data for Marketing Intelligence and Operation Management/ Security and Privacy on Blockchain/ Distributed Storage and Blockchain Technologies for Big Data	Sotirios Tasoulis/ Wutao Wei/ Pan Li, Kim-Kwang Raymond Choo, and Xiaodong Lin/ Hui Li and Han Wang	Beaudry A
	Workshop: Scalable Cloud Data Management (SCDM)	Felix Gessert, Wolfram Wingerath, and Norbert Ritter	Beaudry B
	Workshop: Analysis of Large-scale Disparate Data	Michael Barton, Simon Su, and Brian Panneton	Palos Verdes
	Workshop: Big Data Analytic for Cybercrime Investigation and Prevention	Andrii Shalaginov, Jan William Johnsen, Ambika Shrestha Chitrakar and Asif Iqbal	San Fernando



	Workshop: Big Data Predictive Maintenance using Artificial Intelligence	Aviv Segev, Rituparna Datta, and Ryan Benton	San Pedro
	Workshop: Energy-Efficient Machine Learning and Big Data Analytics	Mohammed Alawad	San Bernardino
	Workshop: Open Science in Big Data (OSBD)	Shannon Quinn, Michael Cotterell, Kyle Johnsen, Nicole Lazar, Suchi Bhandarkar, and John Miller	Silver Lake
	Workshop: Deep Graph Learning: Methodologies and Applications (DGLMA)	Lingfei Wu, Liang Zhao, Jiliang Tang, and Tyler Derr	Echo Park
10: 10 - 10: 30 am Location:	<b>Coffee Break</b> California Foyer		
12: 00 - 1: 30 pm Location:	<b>Lunch (on your own)</b>		
Time	Sessions/Workshops	Session Chair	Location
1: 30 - 3: 30	Tutorial 8: How to build and run a big data platform in the 21st century	Ali Dasdan, Dhruva Borthakur	San Gabriel C
Afternoon workshop (any time between 1:30 am - 7 pm, please check each individual workshop schedule)	Workshop: Applications of Big Data Technology in the Transport Industry	John Easton	Beaudry A
	Workshop: Big Data Analytics in Supply Chains and Transportation	Allan Nengsheng Zhang and Satish Ukkusuri	Beaudry B
	Workshop: Applications of Artificial Intelligence in the Legal Industry	Jianping Zhang, Nathaniel Huber-Fliflet, Robert Keeling, Christian J. Mahoney, and Haozhen Zhao	San Fernando
	Workshop: Big Data Tools, Methods, and Use Cases for Innovative Scientific Discovery (BTSD)	Sangkeun Lee and Travis Johnston	San Pedro
	Workshop: Big Data Analytics for Cyber Intelligence and Defense (BDA4CID)	Stephen McGough	Los Feliz
	Workshop: Big Data Analytic Technology for Bioinformatics and Health Informatics (KDDBHI)	Donghui Wu and Xin Deng	San Bernardino
	Workshop: Graph Techniques for Adversarial Activity Analytics (GTA3)	Jiejun Xu and Hanghang Tong	Silver Lake
	Big Data for Economic and Business Forecasting	Wei Shang, Matthew Harding, Xingfen Wang, Taoyang Wu, Yongxin Tong, and Wei Xu	Echo Park
	Big Data Transfer Learning (BDTL) -- Heterogeneous Representation and Networks	Ming Shao	Mt Washington
3: 40 - 4: 00 pm Location:	<b>Coffee Break</b> California Foyer		

Day 2: Tuesday, December 10, 2019			
7: 40 - 6:00 pm Location:	<b>Registration</b> California Foyer		
8: 30 - 8: 45 am Location:	<b>Opening and Welcome</b> Conf Chairs, PC Chairs, I&G Chair San Francisco/San Jose, Sacramento		
8: 45 - 9: 45 am Location:	<b>Keynote: Prof. Lise Getoor</b> <b>Session Chair: Chaitanya Baru</b> San Francisco/San Jose, Sacramento		
9: 45 - 10: 05 am Location:	<b>Coffee Break</b> California Foyer		
	<b>Poster Session (Set up)</b> San Diego		
Time	Sessions/Workshops	Session Chair	Location
Whole Day Workshops/Special Session (any time between 8 am - 7 pm, please check each individual workshop program schedule)	Special Session: 5th Special Session on Intelligent Data Mining	Uraz Yavanoglu	Palos Verdes
	Workshop: Performance Engineering with Advances in Software and Hardware for Big Data Science (PEASH)	Hui Zhang, Weijia Xu, and Hongfeng Yu	San Pedro
	Workshop: Internet of Things Data Analytics (IOTDA)	Eyhab Al-Masri and Yan Bai	San Bernardino
	Workshop: IoT Big Data and Blockchain	Huaglori Tianfield, Feng Qian	Los Feliz
	Workshop: 6th International Workshop on Privacy and Security of Big Data (PSBD 2019)	Alfredo Cuzzocrea	Mt Washington
	Brain Data Bank Challenge	N. Nan Chu	San Fernando
Morning workshop (any time between 8 am - 1 pm, please check each individual workshop schedule)	Special Session: 2nd Special Session on HealthCare Data	Sultan Turhan, Ozgun Pinarer	Silver Lake
Time	Sessions/Workshops	Session Chair	Location
10: 05 - 12: 10 pm	L1 Novel theoretical models for big data	Wenqing Hu, Missouri S&T	Santa Anita A + B
	L3 high performance/parallel computing platforms for big data	Hong-Linh Truong, Aalto University	Santa Anita C
	L11 Big Data Search Architectures, Scalability and Efficiency	Dawei Zhou, UIUC	San Gabriel A +B
	L20 Cloud/Grid Data Mining-Big Velocity Data	Aibek Musaev, University of Alabama	Santa Barbara A + B
	L23 Algorithms and Systems for Big Data Search	Weijia Xu, University of Texas at Austin	Beaudry A
	Tutorial 6: Large scale semantic graph data management and analytics	Olivier Cure	Beaudry B
	I&G Session 1: Big Data Algorithms & Systems (1)	Petros Zerefos, IBM Research	Santa Barbara C
12: 10 - 2: 00 pm Location:	<b>Lunch (provided by Conference)</b> San Francisco/San Jose, Sacramento		
2: 00 - 3: 00 pm Location:	<b>Keynote: Prof. Yang Qiang</b> <b>Session Chair: Luke Huan</b> San Francisco/San Jose, Sacramento		
3: 00 - 4: 00 pm Location:	<b>Keynote: Prof. Judea Pearl</b> <b>Session Chair: Carlo Zaniolo</b> San Francisco/San Jose, Sacramento		
Time	Sessions/Workshops	Session Chair	Location

Afternoon (any time between 2 pm - 7 pm, please check each individual workshop program schedule)	Workshop: Streaming Systems and Real-Time Machine Learning (STREAM-ML)	Judy Qiu, Geoffrey Fox and Madhav Marathe	Echo Park
4: 00 - 4: 20 pm Location:	<b>Coffee Break</b> California Foyer  <b>Poster Session Sets Up and Displays</b> San Diego		
Time	Sessions/Workshops	Session Chair	Location
4: 20 - 6: 20 pm	L7 cloud/grid/stream computing for big data	Si Zhang, UIUC	Santa Anita A + B
	S5 Distributed, and Peer-to-peer Search	Amr Magdy, UC Riverside	Santa Anita C
	S13 7.a. Complex Big Data Applications in Science, Engineering, Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication		San Gabriel A +B
	S14 7.b. Complex Big Data Applications in Science, Engineering, Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication	Aki-Hiro Sato, Yokohama City University	Santa Barbara A + B
	Tutorial 5: An Overview of the Big Data Approaches for Profitable Social Network Analysis	Elio Masciari, Domenico Saccà	Beaudry A
	I&G Session 2: Big Data & Machine Learning (1)	Su Won Bae, Mobilewalla	Beaudry B
	Special Track on Federated Machine Learning	Yang Liu, Han Yu	Santa Barbara C

Day 3: Wednesday, December 11, 2019			
8: 00 - 6: 00 pm Location:	<b>Registration</b> California Foyer		
8: 45 - 9: 45 am Location:	<b>Keynote: Prof. Ling Liu</b> <b>Session Chair: Latifur Khan</b> San Francisco/San Jose, Sacramento		
9: 45 - 10: 05 am Location:	<b>Coffee Break</b> California Foyer <b>Poster Session Displays</b> San Diego		
Time	Sessions/Workshops	Session Chair	Location
10: 05 - 12: 10 pm	L2 New computational models for big data	Sang-Woo Jun, UC Irvine	Santa Anita A + B
	L4 Search and Mining of variety of data including scientific and engineering, social, sensor/IoT/IoE, and multimedia data	Chao Lan, University of Wyoming	Santa Anita C
	L12 Large-scale Recommendation Systems and Social Media Systems	Aibek Musaev, University of Alabama	San Gabriel A +B
	L17 Complex Big Data Applications in Science, Engineering, Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication	Mohammad Al Hasan, IUPUI	Santa Barbara A + B
	L21 Link and Graph Mining	Yang Zhou, Auburn University	Beaudry A
	Tutorial 2: Taming Unstructured Big Data: Automated Information Extraction from Massive Text	Xuan Wang, Yu Zhang, Qi Li, Jiawei Han	Beaudry B
	I&G Session 3: Big Data & Machine Learning (2)	Baoxu Shi, LinkedIn	Santa Barbara C
	Special session: Information Granulation in Data Science and Scalable Computing	Shusaku Tsumoto, Dominik Slezak, Tzung-Pei Hong, S. L. Wang, Weiping Ding	Los Feliz
Time	Sessions/Workshops	Session Chair	Location
Whole Day Workshops/Special Session (any time between 8 am - 7 pm, please check each individual workshop program schedule)	Workshop: Computational Archival Science: digital records in the age of big data	Mark Hedges, Richard Marciano, and Victoria Lemieux	San Fernando
	Workshop: Big Data Analytics for Cyber Threat Hunting (CyberHunt 2019)	Vasileios Mavroeidis	San Pedro
	Workshop: Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD)	Zhiyuan Chen, Jianwu Wang, Feng Chen, and Yiming Ying	San Bernardino
Morning Workshops/Special Session (any time between 8am-1pm, pls check each individual workshop schedule)	5th Special Session on Intelligent Data Minin	Uraz Yavanoglu	Palos Verdes
	Workshop: Machine Learning for Big Data Analytics in Remote Sensing	Maryam Rahnemoonfar	Mt Washington
12: 10 - 1: 30 pm Location:	<b>Lunch (provided by Conference)</b> San Francisco/San Jose, Sacramento		
1: 30 - 2: 30 pm Location:	<b>Keynote: Dr. Ramanathan Guha</b> <b>Session Chair: Chaitanya Baru</b> San Francisco/San Jose, Sacramento		
Time	Sessions/Workshops	Session Chair	Location
2: 30 - 4: 10 pm	L5 New computational models for big data2	Aparna Varde, Montclair University	Santa Anita A + B

	L6 New computational models for big data3	Ajitesh Srivastava, University of Southern California	Santa Anita C
	L8 autonomic computing and cyber-infrastructure, system architectures, design and deployment	Sang-Woo Jun, UC Irvine	San Gabriel A +B
	L22 Mobility and Big Data	Feng Yu, Youngstown State University, Ohio, USA	Santa Barbara A + B
	I&G Session 4: Big Data Platforms & Frameworks (1)	Dan Goldwasser, Purdue Univ.	Beaudry A
	Special session: Information Granulation in Data Science and Scalable Computing	Shusaku Tsumoto, Dominik Slezak, Tzung-Pei Hong, S. L. Wang, Weiping Ding	Los Feliz
	Panel: Addressing Big Data Heterogeneity Challenges: Recent Advances and Challenges	Vijay Raghavan	Beaudry B
Afternoon workshops (any time between 1:30 pm - 6 pm, please check each individual workshop program schedule)	Workshop: Big Data Engineering and Analytics in Cyber-Physical Systems (BigEACPS)	Akbar Namin	Mt Washington
4: 10 - 4: 30 pm Location:	<b>Coffee Break</b> California Foyer  <b>Poster Session Displays</b> San Diego		
<b>Time</b>	<b>Sessions/Workshops</b>	<b>Session Chair</b>	<b>Location</b>
4: 30 - 6: 15 pm	S2 1.b. New Computational Models for Big Data	Chao Lan, University of Wyoming	Santa Anita A + B
	S3 a. Cloud/Grid/Stream Computing for Big Data	Ariyam Das, UCLA	Santa Anita C
	S6 4.a. Social Web Search and Mining	Mohammad Al Hasan, IUPUI	San Gabriel A +B
	S7 4.c. Algorithms and Systems for Big Data Search	Xuan Wang, UIUC	Santa Barbara A + B
	S15 7.c. Big Data Analytics in Government, Public Sector and Society in General	Rajeev Agrawa, U.S. Army	Beaudry A
	Tutorial 3: Secure and Privacy-Preserving Big-Data Processing	Anton Burtsev, Sharad Mehrotra, Shantanu Sharma	Beaudry B
	I&G Session 5: Big Data Applications: Deep Learning (1)	Fahad Alhasoun, MIT	Santa Barbara C
	Special session: Information Granulation in Data Science and Scalable Computing	Shusaku Tsumoto, Dominik Slezak, Tzung-Pei Hong, S. L. Wang, Weiping Ding	Los Feliz
7: 00 - 9: 00 pm Location	<b>Banquet (Ticket required)</b> San Francisco/San Jose, Sacramento 1. Best Paper/Best Application Paper/Best Student Papers Awards (PC Chairs, I&G Program Chair) 2. IEEE Brain Data Bank Challenges and Competitions (Chair: N. Nan Chu)		

Day 4: Thursday, December 12, 2019			
8: 00 - 3: 00 pm Location:	<b>Registration</b> California Foyer		
Time	Sessions/Workshops	Session Chair	Location
8: 45 - 10: 25 am	L13 software systems to support big data computing	Feng Yu, Youngstown State University, Ohio, USA	Santa Anita A + B
	L14 Privacy Preserving Big Data Collection/Analytics	Anna Squicciarini, Pennsylvania State University	Santa Anita C
	L15 Complex Big Data Applications in Science, Engineering, Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication	Chen Li, UCI	San Gabriel A +B
	L16 Complex Big Data Applications in Science, Engineering, Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication	Yanfeng Fanny Ye, West Virginia University	Santa Barbara A + B
	I&G Session 6: Big Data Platforms & Frameworks (2)	Pranjul Yadav, Criteo	Beaudry A
10: 25 - 10: 45 am Location:	<b>Coffee Break</b> California Foyer		
10: 45 - 12: 15 pm Location:	S9 4.k. Link and Graph Mining	Tim Weninger, University of Notre Dame, USA	Santa Anita A + B
	S10 4.l. Semantic-based Data Mining and Data Pre-processing	Esteban Guillen, Sandia/UMN	Santa Anita C
	S11 4.m. Mobility and Big Data	Ahmed El-Kishky, UIUC	San Gabriel A +B
	Tutorial 4: NewSQL: principles, systems and current trends	Patrick Valduriez, Ricardo Jimenez-Peris	Santa Barbara A + B
	I&G Session 7: Big Data Applications (1)	Geert Janssen, IBM Research	Beaudry A
Time	Sessions/Workshops	Session Chair	Location
Morning Workshops/Special sessions (any time between 8am-1pm, pls check each individual workshop program schedule)	Workshop: Policy-based Autonomic Data Governance (PADG)	Seraphin Calo, Elisa Bertino, and Dinesh Verma	San Fernando
	Workshop: High Performance Big Graph Data Management, Analysis, and Mining (BigGraphs 2019)	Mohammad Hasan, Kamesh Madduri, Nesreen Ahmed, and Shaikh Arifuzzaman	San Pedro
	Workshop: Solar & Stellar Astronomy Big Data (SABiD)	Rafal Angryk, Piet Martens, and Russel White	Los Feliz
	Workshop: Big Data for CyberSecurity (BigCyber)	Karuna Joshi and Sudip Mittal	San Bernardino
	Bigdata Cup Challenges: Suspicious Network Event Recognition	Dominik Slezak, Agnieszka Chadzynska-Krasowska, Joel Holland, Andrzej Janusz, Daniel Kaluza, Bartek Konarski, and Agnieszka Sochal	Santa Barbara C
12: 15 - 1: 30 pm Location:	<b>Lunch (provided by the conference)</b> San Francisco/San Jose, Sacramento		
Time	Sessions/Workshops	Session Chair	Location
1: 30 - 3: 35 pm Location:	L9 Social Web Search and Mining	Manoj Reddy, UCLA	Santa Anita A + B
	L10 Algorithms and Systems for Big Data Search	Karuna Joshi, UMBC	Santa Anita C
	L18 Complex Big Data Applications in Science, Engineering, Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication		San Gabriel A +B
	L19 Techniques and models for transparency/ Interpretability/security/privacy	Bo Dong UT Dallas	Santa Barbara A + B

	I&G Session 8: Big Data Applications (2)	Masha Gorkovenko, Lenovo	Beaudry A
	NSF REU special Symposium	George Mohler, Mohamamd Al Hasan	Beaudry B
3: 35 - 3: 55 pm Location:	<b>Coffee Break</b> California Foyer		
<b>Time</b>	<b>Sessions/Workshops</b>	<b>Session Chair</b>	<b>Location</b>
3: 55 - 5: 40 pm	S1 1.a. Novel Theoretical Models for Big Data		Santa Anita A + B
	S4 3.a. Search and Mining of variety of data including scientific and engineering, social, sensor/IoT/IoE, and multimedia data	Ajitesh Srivastava, University of Southern California	Santa Anita C
	S8 4.h. Computational Modeling and Data Integration	Li Sun, BUPT	San Gabriel A +B
	S12 5.a. Techniques and models for fairness and diversity	Chao Lan, University of Wyoming	Santa Barbara A + B
	Tutorial 9: Deep Learning on Big Data with Multi-Node GPU Jobs	Thomas Breuel, Alex Aizman	Beaudry A
<b>Time</b>	<b>Sessions/Workshops</b>	<b>Session Chair</b>	<b>Location</b>
Afternoon workshops/Special sessions (any time between 1:30 pm - 6 pm, please check each individual workshop program schedule)	NSF REU special Symposium	George Mohler, Mohamamd Al Hasan	Beaudry B
	Workshop: Methodologies to Improve Managing Big Data projects	Jeffrey Saltz	San Fernando
	Workshop: Big data for Intelligent Transportation Systems (BITS)	Steven Jones and Aibek Musaev	San Pedro
	Workshop: Big Media Dataset Construction, Management and Applications	Mingli Song, Mingyu You, Shengcai Liao, Cheng Jin, Yezhou Yang, and Haimiao Hu	Los Feliz
	Workshop: The next frontier of big data from LIDAR	Wang Zhou and Hendrik Hamann	San Bernardino
	Bigdata Cup Challenges: Suspicious Network Event Recognition	Dominik Slezak, Agnieszka Chadzynska-Krasowska, Joel Holland, Andrzej Janusz, Daniel Kaluza, Bartek Konarski, and Agnieszka Sochal	Santa Barbara C

# Keynote Lectures

## Keynote: Responsible Data Science

### *Speaker:*

Lise Getoor, Professor in Computer Science Department, Director of the UC Santa Cruz D3 Data Science Center, University of California, Santa Cruz, USA

### *Abstract:*

Data science is an emerging discipline that offers both promise and peril. Responsible data science refers to efforts that address both the technical and societal issues in emerging data-driven technologies. How can data-driven systems reason effectively about complex dependencies and uncertainty? Furthermore, how do we understand the ethical and societal issues involved in data-driven decision-making? There is a pressing need to integrate algorithmic and statistical principles, social science theories, and basic humanist concepts so that we can think critically and constructively about the socio-technical systems we are building. In this talk, I will overview this emerging area.

### *Short Bio:*

**Lise Getoor** is a professor in the Computer Science Department at UC Santa Cruz and founding director of the Data, Discovery and Decisions (D3) Data Science Research Center at the University of California, Santa Cruz. Her research areas include machine learning, data integration and reasoning under uncertainty, with an emphasis on graph and network data. She has over 250 publications, including 13 best paper awards. She is a Fellow of the Association for Artificial Intelligence, has served as an elected board member of the International Machine Learning Society and the Computing Research Association (CRA). She received her PhD from Stanford University in 2001, her MS from UC Berkeley, and her BS from UC Santa Barbara, and was a professor at the University of Maryland, College Park from 2001-2013.

## Keynote: DataCommons

### *Speaker:*

Ramanathan Guha, Founder and Lead, DataCommons.org, Google, USA

### *Abstract:*

Publicly available data from open sources are a vital resource for students and researchers in a variety of disciplines. Unfortunately, processing these datasets to make them useful --- scraping, cleaning, normalizing, joining --- is tedious, error prone and has to be repeated by every group. DataCommons attempts to alleviate some of this pain by synthesizing a single Knowledge Graph from many different data sources. It links references to the same entities (such as cities, counties, organizations, etc.) across different datasets to nodes on the graph, so that users can access data about a particular entity aggregated from different sources. Like the Web, the DataCommons graph is open - any user can contribute data or build applications powered by the graph. In the Google DataCommons, we are jump-starting the graph with data from publicly available sources such as CDC, Census, BLS, FBI, etc. and are looking to engage with the academic community to take it further.

### *Short Bio:*

**Ramanathan Guha** is the founder and lead for DataCommons.org, a platform which synthesizes a wide range of data sets into a single knowledge graph, for use by students and researchers. He is the creator of widely used web standards such as RSS, RDF and Schema.org, and products such as Google Custom Search, and co-founder of Epinions.com and Alpiri. He is currently a Google Fellow and Vice President at Google. He has a Ph.D. in Computer Science from Stanford University, a Master of Science from University of California, Berkeley and a Bachelor of Technology in Mechanical Engineering from IIT Chennai.

## Keynote: Deception, Robustness and Trust in Big Data Fueled Deep Learning Systems

### *Speaker:*

Ling Liu, Professor, School of Computer Science, Georgia Institute of Technology, USA

### *Abstract:*

We are entering an exciting era where human intelligence is being enhanced by machine intelligence through big data fueled artificial intelligence (AI) and machine learning (ML). However, recent work shows that DNN models trained privately are vulnerable to adversarial inputs. Such adversarial inputs inject small amount of perturbations to the input data to fool machine learning models to misbehave, turning a deep neural network against itself. As new defense methods are proposed, more sophisticated attack algorithms are surfaced. This arms race has been ongoing since the rise of adversarial machine learning. This keynote provides a comprehensive analysis and characterization of the most representative attacks and their defenses. As more and more mission critical systems are incorporating machine learning and AI as an essential component in their real-world big data applications and their big data service provisioning platforms or products, understanding and ensuring the verifiable robustness of deep learning becomes a pressing challenge in the presence of adversarial attacks. This includes (1) the development of formal metrics to quantitatively evaluate and measure the robustness of a DNN prediction with respect of intentional and unintentional artifacts and deceptions, (2) the comprehensive



understanding of the blind spots and the invariants in the DNN trained models and the DNN training process, and (3) the statistical measurement of trust and distrust that we can place on a deep learning algorithm to perform reliably and truthfully. In this keynote talk, I will use empirical analysis and evaluation of our cross-layer strategic teaming defense framework and techniques to illustrate the feasibility of ensuring robust deep learning.

**Short Bio:**

**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. Prof. Liu is an internationally recognized expert in the areas of Big Data Systems and Analytics, Distributed Systems, Database and Storage Systems, Internet Computing, Privacy, Security and Trust. Prof. Liu has published over 300 international journal and conference articles and is a recipient of the best paper award from a number 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, IEEE Edge 2017. Prof. Liu is an elected IEEE Fellow and a recipient of IEEE Computer Society Technical Achievement Award. Prof. Liu has served as general chair and PC chairs of numerous IEEE and ACM conferences in the fields of big data, cloud computing, data engineering, distributed computing, very large databases, World Wide Web, and served as the editor in chief of IEEE Transactions on Services Computing from 2013-2016. Currently Prof. Liu is co-PC chair of The Web 2019 (WWW 2019) and the Editor in Chief of ACM Transactions on Internet Technology (TOIT). Prof. Liu's research is primarily sponsored by NSF, IBM and Intel.

**Keynote: The new science of cause and effect, with reflections on data science and artificial intelligence**

**Speaker:**

Judea Pearl, Chancellor Professor, Departments of Computer Science and Statistics, University of California, Los Angeles, USA

**Abstract:**

The past three decades have seen the development of powerful tools for modeling and computing causal relationships which may have major impact on data science. My talk will illustrate how these tools work in seven tasks:

1. Encoding causal assumptions in transparent and testable way
2. Predicting the effects of actions and policies
3. Computing counterfactuals and finding causes of effects
4. Computing direct and indirect effects (Mediation)
5. Integrating data from diverse sources.
6. Recovering from missing data
7. Discovering causal relations from data

A friendly, nontechnical account of these ideas is available in: "The Book of Why: the new science of cause and effect," Judea Pearl and Dana MacKenzie, (Basic Books, 2018). <http://bayes.cs.ucla.edu/WHY/>

**Short Bio:**

**Judea Pearl** is Chancellor professor of computer science and statistics at UCLA, where he directs the Cognitive Systems Laboratory and conducts research in artificial intelligence, human reasoning, and philosophy of science. He has authored hundreds of research papers and three books: Heuristics (1983), Probabilistic Reasoning (1988) and Causality (2000, 2009) which won of the London School of Economics Lakatos Award in 2002. More recently, he co-authored Causal Inference in Statistics (2016, with M. Glymour and N. Jewell) and "The Book of Why" (2018, with Dana Mackenzie) which introduces causal analysis to a general audience. Pearl is a member of the National Academy of Sciences the National Academy of Engineering, a fellow of the IEEE, the Cognitive Science Society and the Association for the Advancement of Artificial Intelligence. In 2012, he won the Technion's Harvey Prize and the ACM Alan Turing Award "for fundamental contribution to artificial intelligence through the development of a calculus for probabilistic and causal reasoning."

**Keynote: Federated Recommendation Systems**

**Speaker:**

Yang Qiang, New Bright Professor of Engineering, Chair Professor and Head of Department of Computer Science and Engineering, Hong Kong University of Science and Technology, China, Chief AI Officer, WeBank

**Abstract:**

Despite its great progress so far, artificial intelligence (AI) is facing a serious challenge in the availability of high-quality Big Data. In many practical applications, data are in the form of isolated islands. Efforts to integrate the data are increasingly difficult partly due to serious concerns over user privacy and data security. The problem is exacerbated by strict government regulations such as Europe's General Data Privacy Regulations (GDPR). In this talk, I will review these challenges and describe efforts to address them in recommendation systems area. In particular, I will give an overview of recent advances in federated learning and then focus on developments of "federated recommendation systems", which aims to build high-performance recommendation systems by bridging data repositories without compromising data security and privacy.

**Short Bio:**

**Yang Qiang** is the Chief AI Officer of WeBank, China's first internet only bank with more than 100 million customers. He is also a chair professor at Computer Science and Engineering Department at Hong Kong University of Science and Technology (HKUST). His research interests include artificial intelligence, machine learning, especially transfer learning and federated learning. He is a fellow of AAAI, ACM, IEEE, AAAS, etc., and the founding Editor in Chief of the ACM Transactions on Intelligent Systems and Technology (ACM TIST) and the founding Editor in Chief of IEEE Transactions on Big Data (IEEE TBD). He received his PhD from the University of Maryland, College Park in 1989 and has taught at the University of Waterloo and Simon Fraser University. He received the ACM SIGKDD Distinguished Service Award in 2017, AAAI Distinguished Applications Award in 2018, Best Paper Award of ACM TiiS in 2017, and the championship of ACM KDDCUP in 2004 and 2005. He is the past President of IJCAI (2017-2019) and an executive council member of AAAI.

# Conference Paper Presentations

## Regular Paper Sessions

L1	Novel theoretical models for big data	
	BigD423 "Infinite Dropout for training Bayesian models from data streams"	Son Nguyen, Tung Nguyen, Linh Ngo, and Khoat Than
	BigD516 "Detecting Model Changes and their Early Signals Using MDL Change Statistics"	So Hirai and Kenji Yamanishi
	BigD621 "SketchyCoreSVD: SketchySVD from Random Subsampling of the Data Matrix"	Chandrajit Bajaj, Yi Wang, and Tianming Wang
	BigD732 "On the Global Convergence of Continuous-Time Stochastic Heavy-Ball Method for Nonconvex Optimization"	Wenqing Hu, Chris Junchi Li, and Xiang Zhou
	BigD338 "HDMF: Hierarchical Data Modeling Framework for Modern Science Data Standards"	Andrew Tritt, Oliver Rübel, Benjamin Dichter, Ryan Ly, Edward Chang, Donghe Kang, Loren Frank, and Kristofer Bouchard
L2	New computational models for big data	
	BigD592 "Uncertainty-Aware Opinion Inference Under Adversarial Attacks"	Adil Alim, Xujiang Zhao, Jin-Hee Cho, and Feng Chen
	BigD624 "A Streaming model for Generalized Rayleigh with extensions to Minimum Noise Fraction"	Soumyajit Gupta and Chandrajit Bajaj
	BigD686 "CTC-Attention based Non-Parametric Inference Modeling for Clinical State Progression"	Riazat Ryan, Handong Zhao, and Ming Shao
	BigD714 "xSVM: Scalable Distributed Kernel Support Vector Machine Training"	Ruchi Shah, Shaoshuai Zhang, Ying Lin, and Panruo Wu
	BigD758 "MindTheStep-AsyncPSGD: Adaptive Asynchronous Parallel Stochastic Gradient Descent"	Karl Bäckström, Marina Papatriantafilou, and Philippos Tsigas
L3	High performance/parallel computing platforms for big data	
	BigD369 "Eirene: Improving Short Job Latency Performance with Coordinated Cold Data Migration and Scheduler-Aware Task Cloning"	Wei Zhou, K. Preston White, and Hongfeng Yu
	BigD377 "MultiLyra: Scalable Distributed Evaluation of Batches of Iterative Graph Queries"	Abbas Mazloumi, Xiaolin Jiang, and Rajiv Gupta
	BigD452 "Progress-based Container Scheduling for Short-lived Applications in a Kubernetes Cluster"	Yuqi Fu, Shaolun Zhang, Jose Terrero, Ying Mao, Guangya Liu, Sheng Li, and Dingwen Tao
	BigD462 "Mechanism Design for An Incentive-aware Blockchain-enabled Federated Learning Platform"	Kentaroh Toyoda and Allan N. Zhang
	BigD499 "Finding Mutual X at WeChat-Scale Social Network in Ten Minutes"	Conghui He, Shijie Sun, Benli Li, Xiaogang Tu, and Donghai Yu
L4	Search and Mining of variety of data including scientific and engineering, social, sensor/IoT/IoE, and multimedia data	
	BigD280 "NetDyna: Mining Networked Coevolving Time Series with Missing Values"	FNU Hairi, Hanghang Tong, and Lei Ying
	BigD635 "Multivariate Long-Term State Forecasting in Cyber-Physical Systems: A Sequence to Sequence Approach"	Nikhil Muralidhar, Sathappan Muthiah, Kiyoshi Nakayama, Naren Ramakrishnan, and Ratnesh Sharma
	BigD687 "Incremental and Adaptive Feature Exploration over Time Series Stream"	Jingwei ZUO, Karine ZEITOUNI, and Yehia TAHER
	BigD458 "NIOBE: An Intelligent I/O Bridging Engine for Complex and Distributed Workflows"	Kun Feng, Hariharan Devarajan, Anthony Kougkas, and Xian-He Sun
	BigD711 "Understanding Spatio-Temporal Urban Processes"	Lais M. A. Rocha, Aline Bessa, Fernando Chirigati, Eugene OFriel, Mirella Moro, and Juliana Freire
L5	New computational models for big data2	
	BigD262 "Subspace Clustering with Active Learning"	Hankui Peng and Nicos Pavlidis

	BigD368 "Parsimonious Morpheme Segmentation with an Application to Enriching Word Embeddings"	Ahmed El-Kishky, Frank Xu, Aston Zhang, and Jiawei Han
	BigD420 "Finding Stable Clustering for Noisy Data via Structure-aware Representation"	Huiyuan Chen and Jing Li
	BigD424 "Explainable Authorship Verification in Social Media via Attention-based Similarity Learning"	Benedikt Boenninghoff, Steffen Hessler, Dorothea Kolossa, and Robert Nickel
L6	New computational models for big data3	
	BigD433 "Mining News Events from Comparable News Corpora: A Multi-Attribute Proximity Network Modeling Approach"	Hyungsul Kim, Ahmed El-Kishky, Xiang Ren, and Jiawei Han
	BigD437 "On Relaxing Failing Queries over RDF Databases"	Wafaa Mebrek, Badran Raddaoui, and Mohamad AlBilani
	BigD569 "Restricted Recurrent Neural Networks"	Enmao Diao, Jie Ding, and Vahid Tarokh
L7	Cloud/grid/stream computing for big data	
	BigD244 "WiSer: A Highly Available HTAP DBMS for IoT Applications"	Ronald Barber, Christian Garcia-Arellano, Ronen Grosman, Guy Lohman, C. Mohan, Rene Muller, Hamid Pirahesh, Vijayshankar Raman, Richard Sidle, Adam Storm, Yuanyuan Tian, Pinar Tozun, and Yingjun Wu
	BigD271 "Novel Online Algorithms for Nonparametric Correlations with Application to Analyze Sensor Data"	Wei Xiao
	BigD492 "A Bicameralism Voting Framework for Combining Knowledge from Clients into Better Prediction"	Yu-Tung Hsieh, Chuan-Yu Lee, Ching-Chi Lin, Pangfeng Liu, and Jan-Jan Wu
	BigD553 "pSPICE: Partial Match Shedding for Complex Event Processing"	Ahmad Slo, Sukanya Bhowmik, Albert Flaig, and Kurt Rothermel
L8	Autonomic computing and cyber-infrastructure, system architectures, design and deployment	
	BigD367 "Elastic Executor Provisioning for Iterative Workloads on Apache Spark"	Donglin Yang, Wei Rang, Dazhao Cheng, Yu Wang, Jiannan Tian, and Dingwen Tao
	BigD614 "Aptimizer - an Extensible, Open-Source Framework for Hyperparameter Tuning"	Jiayi Liu, Samarth Tripathi, Unmesh Kurup, and Mohak Shah
	BigD626 "TAZeR: Hiding the Cost of Remote I/O in Distributed Scientific Workflows"	Joshua Suetterlein, Ryan Friese, Nathan Tallent, and Malachi Schram
L9	BigD766 "Rapid Prototyping of Hierarchical Agglomerative Clustering Algorithms for Distributed Systems"	Saiyedul Islam, Navneet Goyal, Sundar Balasubramaniam, Poonam Goyal, Achal Agarwal, Kirti Singh Rathore, and Nischay Singh
	Social Web Search and Mining	
	BigD392 "DeepBlue: Bi-layered LSTM for tweet popularity Estimation"	Jian Wen, Zhongbao Zhang, Zichang Yin, Li Sun, Sen Su, and Philip S. Yu
	BigD403 "Predicting Influence Probabilities using Graph Convolutional Networks"	Jing Liu, Yudi Chen, Duanshun Li, Noseong Park, Kisung Lee, and Dongwon Lee
	BigD474 "RecANt: Network-based Recruitment for Active Fake News Correction"	Ajitesh Srivastava, Rajgopal Kannan, Charalampos Chelmiss, and Viktor K. Prasanna
	BigD558 "Modelling Online Comment Threads from their Start"	Rachel Krohn and Tim Weninger
L10	BigD301 "Improving Scalability of Parallel CNN Training by Adjusting Mini-Batch Size at Run-Time"	Sunwoo Lee, Qiao Kang, Sandeep Madireddy, Prasanna Balaprakash, Ankit Agrawal, Alok Choudhary, Richard Archibald, and Wei-keng Liao
	Algorithms and Systems for Big Data Search	
	BigD303 "CS*: Approximate Query Processing on Big Data using Scalable Join Correlated Sample Synopses"	Feng Yu and Wen-Chi Hou
	BigD587 "Clustered Hierarchical Entropy-Scaling Search of Astronomical and Biological Data"	Najib Ishaq, George Student, and Noah Daniels
	BigD646 "Fast Computation of Persistent Homology with Data Reduction and Data Partitioning"	Nicholas O. Malott and Philip A. Wilsey
	BigD703 "Towards Real Time Team Optimization"	Qinghai Zhou, Liangyue Li, and Hanghang Tong

	BigD754 "High Dimensional Data Clustering by means of Distributed Dirichlet Process Mixture Models"	Khadidja Meguelati, Benedicte Fontez, Nadine Hilgert, and Florent Masseglia
L11	Big Data Search Architectures, Scalability and Efficiency	
	BigD380 "b-Bit Sketch Trie: Scalable Similarity Search on Integer Sketches"	Shunsuke Kanda and Yasuo Tabei
	BigD302 "Multi-layer Facial Representation Learning for Early Prediction of Septic Shock"	Chen Lin, Julie Ivy, and Min Chi
	BigD554 "Multi-step Forecasting via Multi-task Learning"	Shayan Jawed, Ahmed Rashed, and Lars Schmidt-Thieme
	BigD620 "Adapted Tree Boosting for Transfer Learning"	Wenjing Fang, Chaochao Chen, Bowen Song, Li Wang, Jun Zhou, and Kenny Zhu
	BigD637 "Collective Tensor Completion with Multiple Heterogeneous Side Information"	Huiyuan Chen and Jing Li
L12	Large-scale Recommendation Systems and Social Media Systems	
	BigD345 "Data Context Adaptation for Accurate Recommendation with Additional Information"	Hyunsik Jeon, Bonhun Koo, and U Kang
	BigD378 "Hybrid Deep Embedding for Recommendations with Dynamic Aspect-Level Explanations"	Huanrui Luo, Ning Yang, and Philip S. Yu
	BigD404 "Effective Utilization of External Knowledge and History Context in Multi-turn Spoken Language Understanding Model"	Yufan Wang, Tingting He, Rui Fan, Wenji Zhou, and Xinhui Tu
	BigD457 "Collaborative Ranking Tags and Items via Cross-domain Recommendation"	Huiyuan Chen and Jing Li
	BigD549 "Sentiment-Aware and Personalized Tour Recommendation"	Prarthana Padia, Kwan Hui Lim, Jeffrey Chan, and Aaron Harwood
L13	Software systems to support big data computing	
	BigD297 "General-Purpose vs. Specialized Data Analytics Systems: A Game of ML & SQL Thrones"	Evdokia Kassela, Nikodimos Provatas, Ioannis Konstantinou, Avrilia Floratou, and Nectarios Koziris
	BigD596 "D3N: A multi-layer cache for the rest of us"	Emine Ugur kaynar, mania abdi, Mohammad Hossein Hajkazemi, Ata Turk, Raja Sambasivan, Larry Rudolph, Peter Desnoyers, and Orran Krieger
	BigD629 "AFrame: Extending DataFrames for Large-Scale Modern Data Analysis"	Phanwadee Sinthong
	BigD671 "News Recommender System Considering Temporal Dynamics and News Taxonomy"	Shaina Raza and Chen Ding
	Privacy Preserving Big Data Collection/Analytics	
L14	BigD586 "OPAL: High performance platform for large-scale privacy-preserving location data analytics"	Axel Oehmichen, Shubham Jain, Andrea Gadotti, and Yves-Alexandre de Montjoye
	BigD770 "SecureGBM: Secure Multi-Party Gradient Boosting"	Zhi Feng, Haoyi Xiong, Chuanyuan Song, Sijia Yang, Baoxin Zhao, Licheng Wang, Zeyu Chen, Liping Liu, and Jun Huan
	BigD332 "Denoising and Verification Cross-Layer Ensemble Against Black-box Adversarial Attacks"	Ka-Ho Chow, Wenqi Wei, Yanzhao Wu, and Ling Liu
	BigD622 "Fast Frequent Pattern Mining without Candidate Generations on GPU by Low Latency Memory Allocation"	Yu-Chen Wu, Mi-Yen Yeh, and Tei-Wei Kuo
	Complex Big Data Applications in Science, Engineering, Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication	
L15	BigD580 "Exascale Deep Learning to Accelerate Cancer Research"	Robert Patton, Travis Johnston, Steven Young, Catherine Schuman, Thomas Potok, Derek Rose, Seung-Hwan Lim, Junghoon Chae, Le Hou, Shahira Abousamra, Dimitris Samaras, and Joel Saltz
	BigD598 "Tale of Three States: Analysis of Large Person-to-Person Online Financial Transactions in Three Baltic Countries"	Rajesh Sharma, Artem Mateush, and Jaan Übi
	BigD601 "TransLand: An Adversarial Transfer Learning Approach for Migratable Urban Land Usage Classification using Remote Sensing"	Yang Zhang, Ruohan Zong, Jun Han, Hao Zheng, Qiuwen Lou, Daniel Zhang, and Dong Wang

L16	Complex Big Data Applications in Science, Engineering, Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication	
	BigD648 "Place Representation Based Bike Demand Prediction"	Yang Zhou and Yan Huang
	BigD677 "Origin-destination Flow Prediction with Vehicle Trajectory Data and Semi-supervised Recurrent Neural Network"	Tao Huang, Yintai Ma, Zhiwei Qin, Jianfeng Zheng, Henry X. Liu, Hongtu Zhu, and Jieping Ye
	BigD715 "Into the Reverie: Exploration of the Dream Market"	Theo Carr, Jun Zhuang, Dwight Sablan, Emma LaRue, Yubao Wu, Mohammad Hasan, and George Mohler
	BigD733 "Virtual Big Data for GAN Based Data Augmentation"	Hadi Mansourifar, Lin Chen, and Weidong Shi
L17	Complex Big Data Applications in Science, Engineering, Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication	
	BigD312 "Crime-GAN: A Context-based Sequence Generative Network for Crime Forecasting with Adversarial Loss"	Guangyin Jin, Qi Wang, Xia Zhao, Yanghe Feng, Qing Cheng, and Jincai Huang
	BigD344 "Kernel-based Multi-Task Contextual Bandits in Cellular Network Configuration"	Xiaoxiao Wang, Xueying Guo, Jie Chuai, Zhitang Chen, and Xin Liu
	BigD349 "MSSTN: Multi-Scale Spatial Temporal Network for Air Pollution Prediction"	Zhiyuan Wu, Yue Wang, and Lin Zhang
	BigD360 "Learning to Differentiate Between Main-articles and Sub-articles in Wikipedia"	Muhao Chen, Changping Meng, Gang Huang, and Carlo Zaniolo
	BigD461 "ABR-HIC: Attention Based Bidirectional RNN for Hierarchical Industry Classification"	Rongzhe Wei, Bo Dong, Kuanzheng Yang, Huan He, Qinghua Zheng, and Jianfei Ruan
L18	Complex Big Data Applications in Science, Engineering, Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication	
	BigD226 "Rail Track Quality and T-Stochastic Neighbor Embedding for Hybrid Track Index"	Ahmed Lasisi, Antonio Merheb, Allan Zarembski, and Nii Attoh-Okine
	BigD274 "DSSLP: A distributed framework for semi-supervised link prediction"	Dalong Zhang, Xianzheng Song, Ziqi Liu, Zhiqiang Zhang, Xin Huang, Lin Wang, and Jun Zhou
	BigD286 "Challenges with Extreme Class-Imbalance and Temporal Coherence: A Study on Solar Flare Data"	Azim Ahmadzadeh, Maxwell Hostetter, Berkay Aydin, Manolis Georgoulis, Dustin Kempton, Sushant Mahajan, and Rafal Angryk
	BigD306 "Regularized Operating Envelope with Interpretability and Implementability Constraints"	Qiyao Wang, Haiyan Wang, Chetan Gupta, and Susumu Serita
	BigD308 "Learning to Discover Curbside Parking Spaces from Vehicle Trajectories"	Yuxin Wen, Jizhou Huang, Chongli Zhu, Miao Fan, and Ying Li
L19	Techniques and models for transparency/interpretability/security/privacy	
	BigD341 "RAPID: Rapid and Precise Interpretable Decision Sets"	Sunny Dhamnani, Dhruv Singal, Ritwik Sinha, M Tharun, and Manish Dash
	BigD570 "Insider Threat Detection via Hierarchical Neural Temporal Point Processes"	Shuhan Yuan, Panpan Zheng, Xintao Wu, and Qinghua Li
	BigD618 "SIM: Open-World Multi-Task Stream Classifier with Integral Similarity Metrics"	Yang Gao, Yi-Fan Li, Bo Dong, Yu Lin, and Latifur Khan
	BigD316 "Differentially Private Robust ADMM for Distributed Machine Learning"	Jiahao Ding, Xinyue Zhang, Mingsong Chen, Kaiping Xue, Chi Zhang, and Miao Pan
	BigD565 "Toward Image Privacy Classification and Spatial Attribution of Private Content"	Haoti Zhong, Anna Squicciarini, David Miller, and Sarah Rajtmajer
L20	Cloud/Grid Data Mining-Big Velocity Data	
	BigD484 "Feature Scoring using Tree-Based Ensembles for Evolving Data Streams"	Heitor Murilo Gomes, Rodrigo Mello, Bernhard Pfahringer, and Albert Bifet
	BigD595 "Would a File by Any Other Name Seem as Malicious?"	Andre Nguyen, Edward Raff, and Aaron Sant-Miller
	BigD266 "JSCN: Joint Spectral Convolutional Network for Cross Domain Recommendation"	Zhiwei Liu, Lei Zheng, Jiawei Zhang, Jiayu Han, and Philip S. Yu



	BigD277 "Origin: Non-Rigid Network Alignment"	Si Zhang, Hanghang Tong, Jiejun Xu, Yifan Hu, and Ross Maciejewski
	BigD352 "Unsupervised Author Disambiguation using Heterogeneous Graph Convolutional Network Embedding"	Ziyue Qiao, Yi Du, Yanjie Fu, Pengfei Wang, and Yuanchun Zhou
L21	Link and Graph Mining	
	BigD372 "Towards Interpretable Graph Modeling with Vertex Replacement Grammars"	Justus Hibshman, Satyaki Sikdar, and Tim Weninger
	BigD486 "Non-local Attention Learning on Large Heterogeneous Information Networks"	Yuxin Xiao, Zecheng Zhang, Carl Yang, and ChengXiang Zhai
	BigD588 "Index Based Efficient Algorithms for Closest Community Search"	Esra Akbas
	BigD756 "Local Search for Group Closeness Maximization on Big Graphs"	Eugenio Angriman, Alexander van der Grinten, and Henning Meyerhenke
	BigD772 "Integrating Local Vertex/Edge Embedding via Deep Matrix Fusion and Siamese Multi-label Classification"	Yang Zhou, Chao Jiang, Zijie Zhang, Dejing Dou, Ruoming Jin, and Pengwei Wang
L22	Mobility and Big Data	
	BigD544 "Scalable Distributed Subtrajectory Clustering"	Panagiotis Tampakakis, Nikos Pelekis, Christos Doukeridis, and Yannis Theodoridis
	BigD705 "DETECT: Deep Trajectory Clustering for Mobility-Behavior Analysis"	Mingxuan Yue, Yaguang Li, Haoze Yang, Ritesh Ahuja, Yao-yi Chiang, and Cyrus Shahabi
	BigD609 "Patient-level Classification on Clinical Note Sequences Guided by Attributed Hierarchical Attention"	Cansu Sen, Thomas Hartvigsen, Xiangnan Kong, and Elke Rundensteiner
	BigD634 "VulnerCheck: A Content-Agnostic Detector for Online Hatred-Vulnerable Videos"	Lanyu Shang, Daniel Zhang, Michael Wang, and Dong Wang
L23	Algorithms and Systems for Big Data Search	
	BigD343 "Experimental Evaluation of Bounded-Depth LSM Merge Policies"	Qizhong Mao, Steven Jacobs, Waleed Amjad, Vagelis Hristidis, Vassilis J. Tsotras, and Neal E. Young
	BigD362 "G-Finder: Approximate Attributed Subgraph Matching"	Lihui Liu, Boxin Du, Jiejun Xu, and Hanghang Tong
	BigD664 "SubGraph2Vec: Highly-Vectorized Tree-like Subgraph Counting"	Langshi Chen, Jiayu Li, Cenk Sahinalp, Madhav Marathe, Anil Vullikanti, Andrey Nikolaev, Egor Smirnov, Ruslan Israfilov, and Judy Qiu
	BigD482 "Semi-supervised Learning over Streaming Data using MOA"	Minh Huong Le Nguyen, Heitor Murilo Gomes, and Albert Bifet
	BigD721 "Generalizing Design of Support Measures for Counting Frequent Patterns in Graphs"	Jinghan Meng, Napath Pitaksiranan, and Yicheng Tu

## Short Paper Sessions

	1.a. Novel Theoretical Models for Big Data	
S1	BigD323 "Sparse Block Regression (SBR) for Big Data with Categorical Variables"	Xiang Liu, Huyunting Huang, Ziyang Tang, Tonglin Zhang, and Baijian Yang
	BigD335 "Discovering Partial Periodic Spatial Patterns in Spatiotemporal Databases"	UDAY KIRAN RAGE and Saideep c
	BigD396 "Hierarchical Demand Forecasting for Factory Production of Perishable Goods"	Can Chen, Yijun Wang, Guoan Huang, and Hui Xiong
	BigD672 "Diagnosis of Neural Network via Backward Deduction"	Peifeng Yin, Lei Huang, Sunhwan Lee, Mu Qiao, Shubhi Asthana, and Taiga Nakamura
	BigD712 "Algorithms on Compressed Time-Evolving Graphs"	Michael Nelson, Sridhar Radhakrishnan, and Chandra Sekharan

	BigD425 "An Interactive Data Quality Test Approach for Constraint Discovery and Fault Detection"	Hajar Homayouni, Sudipto Ghosh, Indrakshi Ray, and Michael Kahn
S2	1.b. New Computational Models for Big Data	
	BigD273 "Subsampled Information Criteria for Bayesian Model Selection in the Big Data Setting"	Lijiang Geng, Yishu Xue, and Guanyu Hu
	BigD408 "A discrete mathematics approach for large scale improvement in classification training time"	Sumedh Yadav and Mathis Bode
	BigD460 "Mixture-based Multiple Imputation Model for Clinical Data with a Temporal Dimension"	Ye Xue, Diego Klabjan, and Yuan Luo
	BigD466 "Ordalia: Deep Learning Hyperparameter Search via Generalization Error Bounds Extrapolation"	Benedetto Buratti and Eli Upfal
	BigD412 "Analysis of Evolutionary Behavior in Self-Learning Media Search Engines"	Nikki Lijing Kuang and Clement H.C. Leung
	BigD682 "Activation Ensembles for Deep Neural Networks"	Diego Klabjan and Mark Harmon
	BigD741 "Online Federated Multitask Learning"	Rui Li, Fenglong Ma, Wenjun Jiang, and Jing Gao
S3	a. Cloud/Grid/Stream Computing for Big Data	
	BigD327 "Investigating Edge vs. Cloud Computing Trade-offs for Stream Processing"	Pedro Silva, Alexandru Costan, and Gabriel Antoniu
	BigD503 "Scaling Out Multi-Way Stream Joins using Optimized, Iterative Probing"	Manuel Dossinger and Sebastian Michel
	BigD708 "Effectively Unified Optimization for Large-scale Graph Community Detection"	Jianping Zeng and Hongfeng Yu
	BigD513 "Approximate querying for the property graph language Cypher"	George Fletcher, Alexandra Poulouvasilis, Petra Selmer, and Peter Wood
	BigD353 "Aves: A Framework for Energy-efficient Stream Analytics across Low-power Devices"	Roshan Bharath Das, Marc X. Makkes, Alexandru Uta, Lin Wang, and Henri Bal
	BigD560 "DataCalc: Ad-hoc Analyses on Heterogeneous Data Sources"	ohannes Luong, Dirk Habich, and Wolfgang Lehner
	BigD272 "Data Services with Bindaas: RESTful Interfaces for Diverse Data Sources"	Pradeeban Kathiravelu, Yusuf Nadir Saghar, Tushar Aggarwal, and Ashish Sharma
S4	3.a. Search and Mining of variety of data including scientific and engineering, social, sensor/IoT/IoE, and multimedia data	
	BigD633 "Cluster-based Computation of Exact Betweenness Centrality in Large Undirected Graphs"	Cecile Daniel, Angelo Furno, and Eugenio Zimeo
	BigD571 "Online Feature Selection with Capricious Streaming Features: A General Framework"	Di Wu, Yi He, Xin Luo, Mingsheng Shang, and Xindong Wu
	BigD774 "Deep Multi-Head Attention Network for Aspect-Based Sentiment Analysis"	Danfeng Yan, Jiyuan Chen, Jianfei Cui, Ao Shan, and Wenting Shi
	BigD355 "Engineering a Parallel $\Delta$ -stepping Algorithm"	Erika Duriakova, Deepak Ajwani, and Neil Hurley
	BigD555 "Co-Representation Learning Framework for the Open-Set Data Classification"	Zhuoyi Wang, Bo Dong, Yu Lin, Md Shihabul Islam, Yigong Wang, and Latifur Khan
	BigD603 "Regression Prediction for Geolocation Aware Through Relative Density Ratio Estimation"	Bo Dong, Jinghui Guo, Zhuoyi Wang, Rong Wu, Yang Gao, and Latifur Khan
	BigD551 "Enhanced Alternate Action Recommender System Using Recurrent Patterns and Fault detection system for Smart Home Users"	Prabhat Mishra, Suresh Kumar Gudla, Amogha D Shanbhag, and Joy Bose
S5	3.c. Distributed, and Peer-to-peer Search	
	BigD358 "A Scalable Algorithm for Multi-class Support Vector Machine on Geo-Distributed Datasets"	Tasnim Kabir and Muhammad Abdullah Adnan
	BigD498 "Aggregate Query Prediction under Dynamic Workloads"	Fotis Savva, Christos Anagnostopoulos, and Peter Triantafillou
	BigD579 "Performance Optimization of SpMV on Spark"	Kun Xie, Che-Rung Lee, and Feng-Yuan Liu
	BigD491 "Fast Record Linkage for Company Entities"	Thomas Gschwind, Christoph Miksovic, Julian Minder, Katsiaryna Mirylenka, and Paolo Scotton



	BigD538 "Learning Data Transformations with Minimal User Effort"	Minh Pham, Craig Knoblock, and Jay Pujara
	BigD730 "IProWA: A Novel Probabilistic Graphical Model for Crowdsourcing Aggregation"	Tianqi Wang, Houping Xiao, Fenglong Ma, and Jing Gao
	BigD314 "GRADI: Towards Group Recommendation Using Attentive Dual Top-Down and Bottom-Up Influences"	Zhixiang He, Chi-Yin Chow, Jia-Dong Zhang, and Ning Li
	BigD451 "SENSE: Semantically Enhanced Node Sequence Embedding"	Swati Rallapalli, Liang Ma, Mudhakar Srivatsa, Ananthram Swami, Heesung Kwong, Graham Bent, and Christopher Simpkin
S6	4.a. Social Web Search and Mining	
	BigD267 "Maximizing Contrasting Opinions in Signed Social Networks"	Kaivalya Rawal and Arijit Khan
	BigD383 "DNA: Dynamic Social Network Alignment"	Li Sun, Zhongbao Zhang, Pengxin Ji, Jian Wen, Sen Su, and Philip S. Yu
	BigD418 "Deep Diffusive Neural Network based Fake News Detection from Heterogeneous Social Networks"	Jiawei Zhang, Bowen Dong, and Philip S. Yu
	BigD495 "Multi-spatial Scale Event Detection from Geo-tagged Tweet Streams via Power-law Verification"	Yi Han, Shanika Karunasekera, Christopher Leckie, and Aaron Harwood
	BigD660 "USTAR: Online Multimodal Embedding for Modeling User-Guided Spatiotemporal Activity"	Amila Silva, Shanika Karunasekera, Christopher Leckie, and Ling Luo
	BigD315 "Improved Table Retrieval Using Multiple Context Embeddings for Attributes"	Mohamed Trabelsi, Brian D. Davison, and Jeffrey Heflin
	BigD402 "Two Problems in Knowledge Graph Embedding: Non-Exclusive Relation Categories and Zero Gradients"	Noseong Park
S7	4.c. Algorithms and Systems for Big Data Search	
	BigD356 "Fast Anomaly Detection in Multiple Multi-Dimensional Data Streams"	Hongyu Sun, Qiang He, Kewen Liao, Timos Sellis, Longkun Guo, Xuyun Zhang, Jun Shen, and Feifei Chen
	BigD384 "Graph matching on social networks without any side information"	Charalampos Davalas, Dimitrios Michail, and Iraklis Varlamis
	BigD388 "Pyramid: A General Framework for Distributed Similarity Search on Large-scale Datasets"	Shiyuan Deng, Xiao Yan, Kelvin K.W. Ng, Chenyu Jiang, and James Cheng
	BigD737 "Adaptive Deep Learning based Time Varying Volume Compression"	Yu Pan, Feiyu Zhu, Tian Gao, and Hongfeng Yu
	BigD406 "Visual Anomaly Detection in Event Sequence Data"	Shunan Guo, Zhuochen Jin, Qing Chen, David Gotz, Hongyuan Zha, and Nan Cao
	BigD611 "Visualization for quality health-care: patient flow exploration."	Veronika Domova and Shiva Sander-Tavallaey
	BigD724 "Interactive Visualization of Time-Varying Hyperspectral Plant Images for High-Throughput Phenotyping"	Feiyu Zhu, Yu Pan, Tian Gao, Harkamal Walia, and Hongfeng Yu
S8	4.h. Computational Modeling and Data Integration	
	BigD340 "OverlapLDA: A Generative Approach for Literature-Based Discovery"	Juncheng Ding and Wei Jin
	BigD670 "Enhancing Itemset Tree Rules and Performance"	Jay Lewis, Ryan Benton, David Bourrie, and Jennifer Lavergne
	BigD379 "D2D-TM: A Cycle VAE-GAN for Multi-Domain Collaborative Filtering"	Linh Nguyen and Tsukasa Ishigaki
	BigD523 "Deep Tensor Factorization for Multi-Criteria Recommender Systems"	Zhengyu Chen, Sibao Gai, and Donglin Wang
	BigD348 "FlexNet: Weakly Supervised Deep Learning Approach in Streaming Environments"	Mahardhika Pratama, Andri Ashfahani, and Abdul Hady
	BigD447 "Task Failure Prediction in Cloud Data Centers Using Deep Learning"	Jiechao Gao, Haoyu Wang, and Haiying Shen
	BigD583 "Arbitrated Dynamic Ensemble with Abstaining and Diversity Forecasting ensembles on Data Streams"	Dihia Boulegane, Albert Bifet, and Giyyarpuram Madhusudan
S9	4.k. Link and Graph Mining	

	BigD283 "Meta-path Reduction with Transition Probability Preserving in Heterogeneous Information Network"	Xiaokai Wei, Zhiwei Liu, Lichao Sun, and Philip S. Yu
	BigD382 "MC2: Unsupervised Multiple Social Network Alignment"	Gen Li, Li Sun, Zhongbao Zhang, Pengxin Ji, Sen Su, and Philip S. Yu
	BigD464 "LATTE: Application Oriented Social Network Embedding"	Lin Meng, Jiyang Bai, and Jiawei Zhang
	BigD533 "BIGMat: A Distributed Affinity-Preserving Random Walk Strategy for Instance Matching on Knowledge Graphs"	Ali Assi, Hamid Mcheick, and Wajdi Dhifli
	BigD539 "Metapath Enhanced Graph Attention Encoder for HINs Representation Learning"	Yuwei Fu, Yun Xiong, Philip S. Yu, Tianyi Tao, and Yangyong Zhu
	BigD593 "Graph Matching via Multi-Scale Heat Diffusion"	Lin Li and Daniel L. Sussman
S10	4.l. Semantic-based Data Mining and Data Pre-processing	
	BigD337 "Hierarchical-Document-Structure-Aware Attention with Adaptive Cost Sensitive Learning for Biomedical Document Classification"	Dandan Fang, Jinyong Zhang, Weizhong Zhao, Xiaowei Xu, Xingpeng Jiang, Xiaohua Hu, and Tingting He
	BigD488 "From Text Classification to Keyphrase Extraction for Short Text"	Song-Eun Lee, Kang-Min Kim, Woo-Jong Ryu, Jemin Park, and SangKeun Lee
	BigD661 "motif2vec: Motif Aware Node Representation Learning for Heterogeneous Networks"	Manoj Dareddy, Mahashweta Das, and Hao Yang
	BigD731 "Community-preserving Graph Convolutions for Structural and Functional Joint Embedding of Brain Networks"	Jiahao Liu, Guixiang Ma, Fei Jiang, Chun-Ta Lu, Philip S. Yu, and Ann Ragin
	BigD768 "The Link Regression Problem"	Bowen Dong, Charu Aggarwal, and Philip S. Yu
	BigD771 "Semi-supervised Classification-based Local Vertex Ranking via Dual Generative Adversarial Nets"	Yang Zhou, Jiaxiang Ren, Sixing Wu, Dejing Dou, Ruoming Jin, Zijie Zhang, and Pengwei Wang
S11	4.m. Mobility and Big Data	
	BigD385 "Study Group Travel Behaviour Patterns from Large-Scale Smart Card Data"	XIANCAI TIAN and BAIHUA ZHENG
	BigD744 "Practical End-to-End Repositioning Algorithm for Managing Bike-Sharing System"	Akihiro Yoshida, Yosuke Yatsushiro, Nozomi Hata, Tatsuru Higurashi, Nariaki Tateiwa, Takashi Wakamatsu, Akira Tanaka, Kenichi Nagamatsu, and Katsuki Fujisawa
	BigD317 "Multimodal, Context-Aware, Feature Representation Learning for Classification and Localization"	Sreyasee Das Bhattacharjee, William J. Tolone, Roy Cherian, and Urmimala Sarkar
	BigD483 "Refining Image Search Results using Multiple Attributes"	Tingting Dong, Shoji Nishimura, and Jianquan Liu
	BigD284 "Gossip Learning: Off the Beaten Path"	Lodovico Giarretta and Sarunas Girdzijauskas
	BigD202 "Big data and traditional Chinese medicine (TCM): what's state of the art?"	David Mainenti
S12	5.a. Techniques and models for fairness and diversity	
	BigD311 "FAE: A Fairness-Aware Ensemble Framework"	Vasileios Iosifidis, Besnik Fetahu, and Eirini Ntoutsi
	BigD591 "FairGAN+: Achieving Fair Data Generation and Classification through Generative Adversarial Nets"	Depeng Xu, Shuhan Yuan, Lu Zhang, and Xintao Wu
	BigD704 "DeepManner: Automatically Determining Manner of Death"	Esteban Guillen, Trilce Estrada and Matthew Cain
	BigD291 "Indirect Adversarial Attacks via Poisoning Neighbors for Graph Convolutional Networks"	Tsubasa Takahashi
	BigD373 "RAMP: Real-Time Anomaly Detection in Scientific Workflows"	J. Dinal Herath, Changxin Bai, Guanhua Yan, Ping Yang, and Shiyong Lu
	BigD456 "Walk2Privacy: Limiting target link privacy disclosure against the adversarial link prediction"	Zhongyuan Jiang, Jianfeng Ma, and Philip S. Yu
	BigD578 "Privacy Bargaining with Fairness: Privacy-Price Negotiation System for Applying Differential Privacy in Data Market Environments"	Kangsoo Jung and Seog Park
S13	7.a. Complex Big Data Applications in Science, Engineering, Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication	

	BigD245 "OD-network-based Pedestrian-path Prediction for People-flow Simulation"	Yu Kitano, Satoshi Kuwamoto, and Akinori Asahara
	BigD255 "Application of Multi-channel 3D-cube Successive Convolution Network for Convective Storm Nowcasting"	Wei Zhang, Lei Han, Juanzhen Sun, Hanyang Guo, and Jie Dai
	BigD295 "TEDM-PU: A Tax Evasion Detection Method Based on Positive and Unlabeled Learning"	Yingchao Wu, Qinghua Zheng, Yuda Gao, Bo Dong, Rongzhe Wei, Fa Zhang, and Huan He
	BigD366 "Attributed Sequence Embedding"	Zhongfang Zhuang, Xiangnan Kong, Elke Rundensteiner, Jihane Zouaoui, and Aditya Arora
	BigD390 "Complex Event Analysis for Traffic Risk Prediction based on 3D-CNN with Mutli-souces Urban Sensing Data"	Ngoc-Thanh Nguyen, Minh-Son Dao, and Koji Zettsu
	BigD394 "A Congestion Diffusion Model with Influence Maximization for Traffic Bottleneck Identification in Metrocity Scales"	Baoxin Zhao, Chengzhong Xu, Siyuan Liu, Juanjuan Zhao, and Li Li
	BigD399 "A King's Ransom for Encryption: Ransomware Classification using Augmented One-Shot Learning and Bayesian Approximation"	Amir Atapour-Abarghouei, Stephen Bonner, and Andrew Stephen McGough
	BigD494 "'Tile & Merge': Distributed Delaunay Triangulations for Cloud Computing"	Laurent Caraffa, Mathieu Brédif, Murat Yirci, and Pooran Memari
S14	7.b	
	BigD563 "Similarity hashing for charged particle tracking"	Sabrina Amrouche, Tobias Golling, Moritz Kiehn, Claudia Plant, and Andreas Salzburger
	BigD564 "Implementing a Domain-Independent Framework to Detect Suspicious Review Patterns with Apache Ecosystem Tools"	Diwen Xue, Willie Yee, and Yueping Wang
	BigD767 "The OTree: Multidimensional Indexing with efficient data Sampling for HPC"	Cesare Cugnasco, Hadrien Calmet, Pol Santamaria, Raül Sirvent, Ane Beatriz Eguzkitza, Guillaume Houzeaux, Yolanda Becerra, Jordi Torres, and Jesus Labarta
	BigD631 "Data Streaming Analysis Framework for 3D non-Cartesian Free-breathing Liver DCE-MRI Perfusion Imaging"	Kun Yang and Pan Li
	BigD642 "Objective Sleep Quality as a Predictor of Mild Cognitive Impairment in Seniors Living Alone"	Brian Chen, Hwee-Pink Tan, and Hwee-Xian Tan
	BigD656 "Scaling Deep Learning Models for Large Spatial Time-Series Forecasting"	Zainab Abbas, Jon Reginbald Ivarsson, Ahmad Al-Shishtawy, and Vladimir Vlassov
	BigD673 "AD2: Improving Quality of IoT Data through Compressive Anomaly Detection"	Aekyeung Moon, Xiaoyan Zhuo, Jialing Zhang, and Seung Woo Son
	BigD675 "Big Data Analytics for Power System Cascading Analysis"	Yousu Chen, Tianzhixi Yin, Renke Huang, Xiaoyuan Fan, and Qihua Huang
S15	7.c. Big Data Analytics in Government, Public Sector and Society in General	
	BigD502 "Validating the Use of Wi-Fi Signals to Estimate Hyperlocal Urban Population"	Nicholas Johnson, Pablo Mandiola, Cyrus Blankinship, Bartosz Bonczak, and Constantine Kontokosta
	BigD706 "Changing labour demands for Data Science and Analytics skills and occupations in Australia"	Nik Dawson, Marian-Andrei Rizoio, Mary-Anne Williams, and Benjamin Johnston
	BigD746 "Modeling human attention by learning from large amount of emotional images"	Macario II Cordel
	BigD755 "Unsupervised Conditional Adversarial Network for Tax Evasion Detection"	Rongzhe Wei, Bo Dong, Xulyu Zhu, Jianfei Ruan, Qinghua Zheng, and Huan He
	BigD254 "Detection of Anomaly State Caused by Unexpected Accident using Data of Smart Card for Public Transportation"	Sakura Yamaki, Shou-de Lin, and Wataru Kameyama
	BigD364 "Ad Blocking Whitelist Prediction for Online Publishers"	Shuai Zhao, Achir Kalra, Chong Wang, Cristian Borcea, and Yi Chen
	BigD623 "STRATUM: A BigData-as-a-Service for Lifecycle Management of IoT Analytics Applications"	Anirban Bhattacharjee, Yogesh Barve, Shweta Khare, Shunxing Bao, Zhuangwei Kang, Aniruddha Gokhale, and Thomas Damiano

## Industry and Government Paper Presentations

I&G Session 1: Big Data Algorithms & Systems (1)		
N209	Scaling Up Heterogeneous Waveform Clustering for Long-Duration Monitoring Signal Acquisition, Analysis, and Interaction: Bridging Big Data Analytics with Measurement Instrument Usage Pattern	Masaharu Goto, Naoki Kobayashi, Gang Ren, and Mitsunori Ogihara
N216	Discovering Dynamic Dependencies from Multivariate Time Series	Xuan-Hong Dang, Syed Yousaf, and Petros Zerfos
N245	Data-Centric Helicopter Failure Anticipation: The MGB Oil Pressure Virtual Sensor Case	Nassia Daouayry, Ammar Mechouche, Pierre-Loic Maisonneuve, Vasile-Marian Scuturici and Jean-Marc Petit
N257	Frequent Causal Pattern Mining: A Computationally Efficient Framework for Estimating Bias-Corrected	Pranjul Yadav, Michael Steinbach, M. Regina Castro, Pedro J. Caraballo, Vipin Kumar, and Gyorgy Simon
N221	Intelligent Exploration of Large-Scale Data: What Can We Learn in Two Passes?	Chandrika Kamath
I&G Session 2: Big Data & Machine Learning (1)		
N207	The Conceptual Background of OPTIMIST's AI Module	Sören Stöhrmann, Vera Kamp, and Reinhard Moratz
N254	MNP Inside Out: A Game Theory Assisted Machine Learning Model to Detect Subscriber Churn Behaviors under China's Mobile Number Portability Policy	Ye Ouyang, Aidong Yang, Shuming Zeng and Fanyu Meng
N242	High Impact Customer Acquisition & Retention Modelling – A Scalable Data Mashup Approach	Kajanan Sangaralingam, Nisha Verma, Aravind Ravi, and Su Won Bae
N218	Demystifying Learning Rate Policies for High Accuracy Training of Deep Neural Networks	Yanzhao Wu, Ling Liu, Juhyun Bae, Ka-Ho Chow, Arun Iyengar, Calton Pu, Wenqi Wei, Lei Yu, and Qi Zhang
I&G Session 3: Big Data & Machine Learning (2)		
N230	Representation Learning in Heterogeneous Professional Social Networks with Ambiguous Social Connections	Baoxu Shi, Jaewon Yang, Tim Weninger, Jing How, and Qi He
N233	DeepCarotene - Job Title Classification with Multi-stream Convolutional Neural Network	Jingya Wang, Kareem Abdelfatah, Mohammed Korayem, and janani balaji
N234	Targeted display advertising: the case of preferential attachment	Saurav Manchanda, Pranjul Yadav, Khoa Doan, and Keerthi Sathiya
N235	A Dynamic Neural Network Model for Click-Through Rate Prediction in Real-Time Bidding	Xianshan Qu, Li Li, Xi Liu, Rui Chen, Yong Ge, and Soo-Hyun Choi
N239	A Framework for Explainable Text Classification in Legal Document Review	Christian J. Mahoney, Jianping Zhang, Nathaniel Huber-fliflet, Peter Gronvall, and Haozhen Zhao
I&G Session 4: Big Data Platforms & Frameworks (1)		
N213	Federated Multimodal Big Data Storage & Analytics Platform for Additive Manufacturing	Kareem Aggour, Vijay Kumar, Paul Cuddihy, Jenny Williams, Vipul Gupta, Laura Dial, Tim Hanlon, Justin Gambone, and Joseph Vinciguerra
N219	ACE-An Anomaly Contribution Explainer for Cyber-Security Applications	Xiao Zhang, Manish Marwah, I-ta Lee, Martin Arlitt, and Dan Goldwasser

N244	InfDetect: a Large Scale Graph-based Fraud Detection System for E-Commerce Insurance	Cen Chen, Chen Liang, Jianbin Lin, Li Wang, Ziqi Liu, Xinxing Yang, Jun Zhou, Yang Shuang, and Yuan Qi
N249 (Short)	Streamlined and Accelerated Cyber Analyst Workflows with RAPIDS	Nicholas Becker, Ayush Dattagupta, Eli Fajardo, Prem Gali, Bianca Rhodes, Bartley Richardson, and Bhargav Suryadevara
I&G Session 5: Big Data Applications: Deep Learning (1)		
N250	An Assistive Learning Workflow on Annotating Images for Object Detection	Vivian Wen Hui Wong, Max Ferguson, Kincho H. Law, and Yung-Tsun Tina Lee
N256	Mining Vessel Trajectories for Illegal Fishing Detection	Amir Yaghoubi Shahir, Mohammad A. Tayebi, Uwe Glässer, Tilemachos Charalampous, Zahra Zohrevand, and Hans Wehn
N210	Visual Analytics for Deep Embeddings of Large Scale Molecular Dynamics Simulations	Junghoon Chae, Debsindhu Bhowmik, Heng Ma, Arvind Ramanathan, and Chad Steed
N238 (Short)	Streetify: Using Street View Imagery and Deep Learning for Urban Streets Development	Fahad Alhasoun and Marta Gonzalez
I&G Session 6: Big Data Platforms & Frameworks (2)		
N208	Vertica Flattened Tables and Live Aggregate Projections: A Column-based Alternative to Materialized Views for Analytics	Yuanzhe Bei, Thao Pham, Akshay Aggarwal, Nga Tran, Jaimin Dave, Chuck Bear, and Michael Leuchtenburg
N220	Forward Index Compression for Instance Retrieval in an Augmented Reality Application	Qi Wang, Michal Siedlaczek, Yen-Yu Chen, Michael Gormish, and Torsten Suel
N227	An Efficient Cloud-Based Framework for Digital Media Knowledge Extraction	Chaitanya Kanchibhotla, Pruthvi Raj Venkatesh, DVLN Somayajulu, and Radha krishna P
N229	Delog: A High-Performance Privacy Preserving Log Filtering Framework	Amey Agrawal, Abishek Dixit, Namrata Shettar, Darshil Kapadia, Vikram Agrawal, Rajat Gupta, and Rohit Karlupia
I&G Session 7: Big Data Applications (1)		
N232	EdgeInsight: Characterizing and Modeling the Performance of Machine Learning Inference on the Edge and Cloud	Philipp Ross and Andre Luckow
N211	Large Data Flow Graphs in Limited GPU Memory	Geert Janssen, Vladimir Zolotov, and Tung D. Le
N215	Contextual Price Features for e-Commerce Search Ranking	Ishita Kamal Khan, Aritra Mandal, and Prathyusha Senthil Kumar
N236 (Short)	Empirical Comparisons of CNN with Other Learning Algorithms for Text Classification in Legal Document Review	Robert Keeling, Rishi Chhatwal, Nathaniel Huber-fliflet, Jianping Zhang, Fusheng Wei, Haozhen Zhao, Shi Ye, and Han Qin
I&G Session 8: Big Data Applications (2)		
N246	Towards Power Efficiency in Deep Learning on Data Center Hardware	Miroslav Hodak, Maria Gorkovenko, and Ajay Dholakia
N247	High-Resolution Road Vehicle Collision Prediction for the City of Montreal	Antoine Hébert, Timothée Guédon, Tristan Glatard, and Brigitte Jaumard
N226 (short)	Deep Transfer Learning for Thermal Dynamics Modeling in Smart Buildings	Zhanhong Jiang and Young M. Lee



# Tutorials

## ***TUTORIAL 1: Process mining: Leveraging event data to understand and improve organizations***

### **Presenters:**

**Henrik Leopold**, han.van.der.aa@hu-berlin.de  
**Han van der Aa**

### **Abstract:**

Process mining is a family of data analysis methods that aims to discover, monitor, and improve organizational processes by analyzing data from so-called event logs. These event logs are generated by various information systems that are used in an organization and, therefore, capture how organizational processes are actually executed. The main difference to traditional data analysis techniques is that process mining explicitly focuses on the process perspective. That is, it aims to reveal the complex order relations among the activities captured in the event log. In this tutorial, we give an introduction into the field of process mining and focus on its two most common tasks: (1) process discovery and (2) conformance checking. In the discovery part, we show that process mining techniques can be used to learn and visualize how a process is actually running in practice. In the conformance checking part, we explain how process mining can be used to detect differences between intended behavior (captured in a normative specification) and actual behavior (as found in the event log). Besides introducing the required theory and mechanisms behind discovery and conformance checking, we will put an emphasis on demonstrating how process discovery and conformance checking can be conducted using the open-source tool ProM. In this way, participants will learn how the introduced concepts can be applied and how they can successfully use process mining themselves.

## ***TUTORIAL 2: Taming Unstructured Big Data: Automated Information Extraction from Massive Text***

### **Presenters:**

**Xuan Wang**, xwang174@illinois.edu  
**Yu Zhang**  
**Qi Li**  
**Jiawei Han**

### **Abstract:**

Text data is a powerful information source that covers almost every aspect of our life. Automated information extraction has attracted considerable attention with various approaches developed to mine structured knowledge from unstructured text. In this tutorial, we present an organized picture of automated information extraction from massive text to answer the need of a systematic review and comparison of the techniques. We first introduce major tasks of information extraction such as named entity recognition and relation extraction. Then we introduce downstream applications such as heterogeneous information network construction and claim mining that utilize the extracted information. Specifically, we focus on the methods that are scalable, effective, minimum supervised and working on various kinds of text (e.g., news and biomedical science). We also demonstrate on a real-world dataset, PubMed that includes over 29 million biomedical literature, how the heterogeneous information network can be constructed and how the scientific claims can be automatically retrieved based on automated information extraction. The covered topics will be interesting to both advanced researchers and beginners in data mining, text mining, natural language processing and machine learning.

## ***TUTORIAL 3: Secure and Privacy-Preserving Big-Data Processing***

### **Presenters:**

**Anton Burtsev**  
**Sharad Mehrotra**  
**Shantanu Sharma**, shantanu.sharma@uci.edu

### **Abstract:**

Over the last decade, public and private clouds emerged as de facto platforms for big-data analytical workloads. Outsourcing one's data to the cloud, however, comes with multiple security and privacy challenges. In a world where service providers can be located anywhere in the world, fall under varying legal jurisdictions, i.e., be a subject of different laws governing privacy and confidentiality of one's data, and be a target of well-sponsored (sometimes even government-sponsored) security attacks protecting data in a cloud is far from trivial. This tutorial focuses on two principal lines of research (cryptographic- and hardware-based) aimed to provide secure processing of big-data in a modern cloud. First, we focus on cryptographic (encryption- and secret-sharing-based) techniques developed over the last two decades and specifically compare them based on efficiency and information leakage. We demonstrate that despite extensive research on cryptography, secure query processing over outsourced data remains an open challenge. We then survey the landscape of emerging secure hardware, i.e.,

recent hardware extensions like Intel's Software Guard Extensions (SGX) aimed to secure third-party computations in the cloud. Unfortunately, despite being designed to provide a secure execution environment, existing SGX implementations suffer from a range of side-channel attacks that require careful software techniques to make them practically secure. Taking SGX as an example, we will discuss representative classes of side-channel attacks, and security challenges involved in the construction of hardware-based data processing systems. We conclude that neither cryptographic techniques nor secure hardware are sufficient alone. To provide efficient and secure large-scale data processing at the cloud, a new line of work that combines software and hardware mechanisms is required. We discuss an orthogonal approach designed around the concept of data partitioning, i.e., splitting the data processing into cryptographically secure and non-secure parts. Finally, we will discuss some open questions in designing secure cryptographic techniques that can process large-sized data efficiently.

#### ***TUTORIAL 4: NewSQL: principles, systems and current trends***

##### **Presenters:**

**Patrick Valduriez**, Patrick.Valduriez@inria.fr  
**Ricardo Jimenez-Peris**

##### **Abstract:**

NewSQL is the latest technology in the big data management landscape, enjoying a fast-growing rate in the DBMS and BI markets. NewSQL combines the scalability and availability of NoSQL with the consistency and usability of SQL. By providing online analytics over operational data, NewSQL opens up new opportunities in many application domains where real-time decision is critical. Important use cases are Google Adwords, proximity marketing, real-time pricing, risk monitoring, real-time fraud detection, etc. NewSQL may also simplify data management, by removing the traditional separation between operational database and data warehouse / data lake (no more ETLs!). However, a hard problem is scaling out transactions in mixed operational and analytical (HTAP) workloads over big data, possibly coming from different data stores (HDFS, SQL, NoSQL). Today, only a few NewSQL systems have solved this problem. This tutorial provides an in-depth presentation of NewSQL, with its principles, architectures and techniques. We provide a taxonomy of NewSQL systems based on major dimensions including targeted workloads, capabilities and implementation techniques. We illustrate with popular NewSQL systems such as Google F1/Spanner, LeanXcale, CockroachDB, SAP HANA, MemSQL and Splice Machine. In particular, we give a spotlight on some of the more advanced systems. We also compare with major NoSQL and SQL systems, and discuss integration within big data ecosystems and corporate information systems. Finally, we discuss the current trends and research directions.

#### ***TUTORIAL 5: An Overview of the Big Data Approaches for Profitable Social Network Analysis***

##### **Presenters:**

**Elio Masciari**, elio.masciari@unina.it  
**Domenico Sacca**

##### **Abstract:**

The pervasive diffusion of social networks caused the generation of unprecedented amounts of heterogeneous data. Thus, traditional approaches quickly became unpractical for real life applications. More in detail, the analysis of user generated data by popular social networks like Facebook, Twitter, Instagram, LinkedIn to cite a few, poses quite intriguing challenges for both research and industry communities for analyzing user behavior, user interactions, link evolution, opinion spreading and several other important tasks. This tutorial will focus on the requirements needed for effective analysis of these new kind of data by analyzing some of the most recent approaches in literature. No specific prerequisites are needed, except the basic notions of graph theory, as we aim at guiding the attendees through a high-level tour of the most recent approaches proposed by both researchers and companies. In particular, we will focus on the Big Data peculiar features of SN by analyzing the best solutions according to state of the art. Moreover, as gathering reliable data for research purposes is crucial, we will explain how to properly get huge datasets.

#### ***TUTORIAL 6: Large scale semantic graph data management and analytics***

##### **Presenters:**

**Olivier Cure**, olivier.cure@u-pem.fr

##### **Abstract:**

After years of research and development, standards and technologies for semantic data are sufficiently mature to be used as the foundation of novel data science projects that employ semantic technologies in various application domains such as bioinformatics, materials science, intelligence, and social science. Typically, such projects are carried out by domain experts who have a conceptual understanding of semantic technologies but lack the expertise to choose and to employ existing data management and analytical solutions for the semantic data in their

project. For such experts, including domain-focused data scientists, business analysts, project coordinators, and project engineers, our tutorial will deliver a practitioner's guide to semantic data management and analytics. We will discuss the following important aspects of graph-based semantic data management and demonstrate how to address these aspects in practice by using mature, production-ready tools: Storing and querying semantic data; automated reasoning; integrating external data and knowledge; and analytics.

## ***TUTORIAL 7: Industrial AI: Machine Learning for Maintenance and Repair***

### **Presenters:**

**Chetan Gupta**

**Ahmed Farahat**, Ahmed.Farahat@hal.hitachi.com

### **Abstract:**

Industrial AI is concerned with the application of Artificial Intelligence (AI), Machine Learning (ML) and related technologies towards addressing real world challenges in industrial and societal domains. These challenges can be categorized into the horizontal areas of maintenance and repair (M&R), quality, operations, safety, etc. and have applications in a large number of verticals. These applications will be profound and far reaching impact over the next several years and decades. One of the key horizontals in Industrial AI is Maintenance and Repair (M&R). This tutorial presents an overview of the application of machine learning for industrial operations with a focus on the M&R of physical equipment. To set the context, we will begin with an overview of the M&R business, and introduce a taxonomy of M&R problems that can be solved using AI & ML. We will then deep dive into recent applications in which new modeling techniques have been introduced to solve unique challenges in the M&R, such as using LSTMs and Functional Neural Networks (FNNs) for addressing prognostics problems, using RL for health indicator learning, GANs for generating failure data, etc. Finally, we will present some open problems in Industrial AI and discuss how the research community can shape the future of the next industrial revolution. We hope that by the end of the tutorial, the attendees will not only have a better appreciation for the space of Industrial AI but will be exposed new real-world problems and cutting-edge solutions.

## ***TUTORIAL 8: How to build and run a big data platform in the 21st century***

### **Presenters:**

**Ali Dasdan**, adasdan@atlassian.com

**Dhruba Borthakur**

### **Abstract:**

We want to show that building and running a big data platform for both streaming and bulk data processing for all kinds of applications involving analytics, data science, reporting, and the like in today's world can be as easy as following a checklist. We live in a fortunate time that many of the components needed are already available in the open source or as a service from commercial vendors. We show how to put these components together in multiple sophistication levels to cover the spectrum from a basic reporting need to a full-fledged operation across geographically distributed regions with business continuity measures in place. We plan to provide enough information and checklists to the audience that this tutorial can also serve as a goto reference in the actual process of building and running.

## ***TUTORIAL 9: Deep Learning on Big Data with Multi-Node GPU Jobs***

### **Presenters:**

**Thomas Breuel**, tbreuel@nvidia.com

**Alex Aizman**

### **Abstract:**

Both traditional machine learning (clustering, decision trees, parametric models, cross-validation, function decompositions) and deep learning (DL) are often used for the analysis of big data on hundreds of nodes (clustered servers). However, the systems and I/O considerations for multi-node deep learning are quite different from traditional machine learning. While traditional machine learning is often well served by MapReduce style infrastructure (Hadoop, Spark), distributed deep learning places different demands on hardware, storage software, and networking infrastructure. In this tutorial, we cover: • the structure and properties of large-scale GPU-based deep learning systems • large-scale distributed stochastic gradient descent and supporting frameworks (PyTorch, TensorFlow, Horovod, NCCL) • common storage and compression formats (TFRecord/tf.Example, DataLoader, etc.) and their interconnects (Ethernet, Infiniband, RDMA, NVLINK) • common storage architectures for large-scale DL (network file systems, distributed file systems, object storage) • batch queuing systems, Kubernetes, and NGC for scheduling and large-scale parallelism • ETL techniques including distributed GPU-based augmentation (DALI) The tutorial will focus on techniques and tools by which deep learning practitioners can take advantage of these technologies and move from single-



desktop training to training models on hundreds of GPUs and petascale datasets. It will also help researchers and system engineers to choose and size the systems necessary for such large-scale deep learning. Participants should have some experience in training deep learning models on a single node. The tutorial will cover both TensorFlow and PyTorch frameworks as well as additional open-source tools required to scale deep learning to multi-node storage and multi-node training.

# Workshops

Computational Archival Science		
Workshop Chairs: Mark Hedges, Victoria Lemieux, Richard Marciano		
Time	Title	Presenter/Author
8:45 – 9:00	Welcome	Mark Hedges (King's College London), Richard Marciano (U. Maryland), Victoria Lemieux (U. British Columbia)
<b>9:00 – 9:40</b>	<b>SESSION 1: Computational Thinking in Archival Science</b>	
9:00 – 9:20	Computational Thinking in Archival Science Research and Education	Bill Underwood
9:20 – 9:40	Reframing Digital Curation Practices through a Computational Thinking Framework	Richard Marciano, Dev Pradhan, Chienxi Liu & Kanishka Jain
<b>9:45 – 10:05</b>	<b>Coffee Break</b>	
<b>10:05 – 11:05</b>	<b>SESSION 2: Archival Thinking in Computational Science</b>	
10:05 – 10:25	An Intelligent Class: The Development off a Novel Context Capturing Framework For The Functional Classification Of Records	Nathaniel Payne
10:25 – 10:45	Extending the Scope of Computational Archival Science: A Case Study on Leveraging Archival and Engineering Approaches to Develop a Framework to Detect and Prevent “Fake Video”	Hoda Hamouda
10:45 – 11:05	ArchContract: using smart contracts for disposition	
<b>11:05 – 11:45</b>	<b>SESSION 3: Media Archives</b>	
11:05 – 11:25	Preliminary Analysis of a Large-Scale Digital Entertainment Development Archive: A Case Study of the Entertainment Technology Center's Projects	Eric Kaltman
11:25 – 11:45	Building the National Radio Recordings Database: A Big Data Approach to Documenting Audio Heritage	Shawn VanCour
<b>11:45 – 12:05</b>	<b>SESSION 4: CAS and the Representation of Objects (Part 1)</b>	
11:45 – 12:05	What Computational Archival Science Can Learn from Art History and Material Culture Studies	Lyneise Williams
<b>12:10 – 1:30</b>	<b>Lunch</b>	
<b>1:30 – 1:50</b>	<b>SESSION 4: CAS and the Representation of Objects (Part 2)</b>	
1:30 – 1:50	Digital Legacies on Paper: Reading Punchcards with Computer Vision	Greg Jansen
<b>1:50 – 2:30</b>	<b>SESSION 5: CAS Architecture</b>	
1:50 – 2:10	Enterprise Architecture – A Value Proposition for Records Professionals	Shadrack Katuu
2:10 – 2:30	Using Data Partitions and Stateless Servers to Scale Up Fedora Repositories	Greg Jansen
<b>2:30 – 3:10</b>	<b>SESSION 6: Knowledge Organization</b>	
2:30 – 2:50	Automated interpretability of linked data ontologies: an evaluation within the cultural heritage domain	Nuno Freire
2:50 – 3:10	Towards a Flexible System Architecture for Automated Knowledge Base Construction Frameworks	Osman Din
<b>3:10 – 4:10</b>	<b>SESSION 7: Open Mic Updates</b>	
3:10 – 3:30	International CAS Network	Mark Hedges
3:30 – 3:50	NARA at the Inflection Point: The Looming 2022 Date for Going Digital And What It Means from an AI Perspective	Jason R. Baron
3:50 – 4:10	Computational Thinking Practices in the Workshop Papers	Bill Underwood
<b>4:10 – 4:30</b>	<b>Coffee Break</b>	
<b>4:30 – 5:00</b>	<b>Closing Remarks</b>	

## International Workshop on Big Data Analytics for Cyber Threat Hunting (CyberHunt 2019)

Wednesday, December 11, 2019  
Location: San Pedro

Workshop Chair: Vasileios Mavroeidis – University of Oslo  
Program Committee Chair: Audun Jøsang – University of Oslo

Time	Title	Presenter/Author
10:05 – 10:10 am	Opening Remarks and Welcome	Vasileios Mavroeidis
10:10 – 10:50 am	<b>Keynote Speech</b> <b>Title: Open Command and Control (OpenC2) – Towards Active Cyber Defense by Standardizing Course of Action Operations</b>	<b>Joe Brule</b> <b>Cyber-Engineer in the Capabilities Directorate – National Security Agency (NSA)</b> <b>Co-Chair for the OASIS OpenC2 TC</b>
10:50 – 11:10 am	Do's and Don'ts of Distributed Intrusion Detection for Industrial Network Topologies	Peter Schneider
11:10 – 11:30 am	A Framework for Cyber Threat Intelligence Extraction from Raw Log Data	Max Landauer, Florian Skopik, Markus Wurzenberger, Wolfgang Hotwagner, and Andreas Rauber
11:30 – 11:50 am	Worth the Wait? Time Window Feature Optimization for Attack Classification	Casey Wilson, Xenia Mountrouidou, and Anna Little
11:50 – 12:10 pm	A Framework for Healthcare Security Practice Analysis, Modeling and Incentivization	Prosper Kandabongee Yeng, Bian Yang, and Einar Arthur Snekkenes
12:10 – 01:30 pm Location:	<b>Lunch</b> San Francisco/San Jose, Sacramento	
1:30 – 2:10 pm	<b>Keynote Speech</b> <b>Title: Ecosystem for Cyber Threat Intelligence</b>	<b>Prof. Audun Jøsang</b> <b>Leader of the Digital Security Research Group – University of Oslo (UiO)</b>
2:10 – 2:30 pm	Detecting Adversary using Windows Digital Artifacts	Seng Pei Liew and Satoshi Ikeda
2:30 – 2:50 pm	Detection of Phishing websites using Generative Adversarial Algorithms	Pierrick Robic--Butez and Thu Yein Win
2:50 – 3:10 pm	Examination of Double Arbiter PUFs on Security against Machine Learning Attacks	Meznah Alamro, Yu Zhuang, Ahmad Aseeri, and Mohammed Alkathiri
3:10 – 3:30 pm	Learning to Generate Diverse and Authentic Reviews via an Encoder-Decoder Model with Transformer and GRU	Kaifu Jin, Xi Zhang, and Jiayuan Zhang
3:30 – 3:50 pm	Automatic Extraction of Personality from Text: Challenges and Opportunities	Nazar Akrami, Johan Fernquist, Tim Isbister, Lisa Kaati, and Björn Pelzer
3:50 – 4:10 pm	Byakko: Automatic Whitelist Generation based on Occurrence Distribution of Features of Network Traffic	Nobuyuki Kanaya, Yu Tsuda, Yuuki Takano, and Daisuke Inoue
4:10 – 4:30 pm	Malicious URL Linkage Analysis and Common Pattern Discovery (Remotely – Video Conference Presentation)	Shin-Ying Huang, Tzu-Hsien Chuang, Shi-Meng Huang, and Tao Ban
4:30 – 4:35 pm	Closing Remarks	Vasileios Mavroeidis

<b>The 2<sup>nd</sup> International Workshop on Big Data engineering and Analytics in Cyber-Physical Systems (BigEACPS'19)</b> Wednesday (Afternoon: 2:30 – 7:00), December 11, 2019 <i>Workshop Chairs: Akbar Siami Namin</i>		
Time	Title	Presenter/Author(s)
2:30 – 3:00	<i>Accelerated Evaluation of Autonomous Drivers using Neural Network Quantile Generators</i>	<b>Edward Schwalb</b>
3:00 – 3:30	<i>Can Machine/Deep Learning Classifiers Detect Zero-Day Malware with High Accuracy?</i>	<b>Faranak Abri</b> , Sima Siami-Namini, Mahdi Adl Khanghah, Fahimeh Mirza Soltani, and Akbar Siami Namin
3:30 – 4:00	<i>HackerNets: Interactive Network Visualization for Media Conversations in Internet of Things, Big Data, and Cybersecurity</i>	Hao Van, Huyen N. Nguyen, and <b>Tommy Dang</b>
4:00 – 4:30	<b>Coffee Break</b>	
4:30 – 5:00	<i>Provenance-aware workflow for data quality management and improvement for large continuous scientific data streams</i>	<b>Jitendra Kumar</b> , Michael Crow, Ranjeet Devarakonda, Michael Giansiracusa, Kavya Guntupally, Joseph V. Olatt, Zach Price, Harold A. Shanafield, and Alka Singh
5:00 – 5:30	<i>Comparison of Support Vector Machine and Gradient Boosting Regression Tree for Predicting Spatially Explicit Life Cycle Global Warming and Eutrophication Impacts: A case study in corn production</i>	XIAOBO Romeiko, zhijian Guo, and <b>Yulei Pang</b>
5:30 – 6:00	<i>Probability Density Representation and Inference using Nadaraya-Watson Estimator</i>	<b>Edward Schwalb</b>

6:00 – 6:30	<i>The Performance of LSTM and BiLSTM in Forecasting Time Series</i>	Sima Siامي-Namini, Neda Tavakoli, and <b>Akbar Siامي Namin</b>
6:30 – 6:55	<i>MTSAD: Multivariate Time Series Abnormality Detection and Visualization</i>	Vung Pham, Ngan Nguyen, Jie Li, Jon Hass, Yong Chen, and <b>Tommy Dang</b>
6:55 – 7:00	<b>Closing Remarks</b>	

<b>PEASH</b> <i>Workshop Chairs: Hui Zhang, Weijia Xu, Hongfeng Yu</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
8:00am – 8:15am	<b>PEASH'19 Opening Remarks</b>	
8:15am – 8:35am	<i>Parallel R Computing on the Web</i>	Ranjini Subramanian
8:35am – 8:55am	<i>An Evaluation of RDMA-based Message Passing Protocols</i>	Shahram Ghandeharizadeh
8:55am – 9:15am	<i>Parallel Training via Computation Graph Transformation</i>	Fei Wang
9:15am – 9:35am	<i>Accelerating RNN on FPGA with Efficient Conversion of High-Level Designs to RTL</i>	Zongze Li
9:45am – 10:05am	<b>Coffee Break</b>	
10:05am – 10:25am	<i>Parallelized Topological Relaxation Algorithm</i>	Guangchen Ruan
10:25am – 10:45am	<i>Transparent In-memory Cache Management in Apache Spark based on Post-Mortem Analysis</i>	Atsuya Nasu
10:45am – 11:05am	<i>A Fast Exact Viewshed Algorithm on GPU</i>	Faisal Qarah
11:05am – 11:25am	<i>Spatial-Temporal Scientific Data Clustering via Deep Convolutional Neural Network</i>	Jianxin Sun
11:25am – 11:45am	<i>A GPU based parallel algorithm for computing the Sparse Fast Fourier Transform (SFFT) of k-sparse signals</i>	Fahad Saeed
12:10pm – 2:00pm	<b>Lunch Break</b>	
2:20pm – 2:40pm	<i>Plant Event Detection from Time-Varying Point Clouds</i>	Tian Gao
2:40pm – 3:00pm	<i>Performance Comparison of Julia Distributed Implementations of Dirichlet Process Mixture Models</i>	Ruizhu Huang
3:00pm – 3:20pm	<i>Parallel Hybrid Metaheuristics with Distributed Intensification and Diversification for Large-scale Optimization in Big Data Statistical Analysis</i>	Wendy Tam
3:20pm – 3:40pm	<i>An "On The Fly" Framework for Efficiently Generating Synthetic Big Data Sets</i>	Karm Mason
3:40pm – 4:00pm	<i>Auto-CNNp: a component-based framework for automating CNN parallelism</i>	Soulaimane GUEDRIA
4:00pm – 4:20pm	<b>Coffee Break</b>	
4:00pm – 4:20pm	<i>Constructing Suffix Array of Next-Generation Sequencing upon In-Memory Lookup Cloud and MapReduce</i>	Meng-Huang Lee
4:20pm – 4:40pm	<i>View Selection in Knot Deformation</i>	Juan Lin
4:40pm – 5:00pm	<b>PEASH'19 Closing Remarks</b>	

<b>5th International Workshop on Methodologies to Improve Big Data Projects</b> <i>Workshop Chair: Jeffrey Saltz</i> <b>December 12<sup>th</sup></b>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
1:30	A Methodology for Cross-Platform, Event-Driven Big Data Analytics-as-a-Service	Claudio Ardagna
1:55	A Framework for Identifying and Prioritizing Data Analytics Opportunities in Additive Manufacturing	Hyunseop Park
2:20	Context-Augmented Software Development Projects: Literature Review and Preliminary Framework	Glaucia Melo
2:45	SKI: An Agile Framework for Data Science	Jeffrey Saltz

3:10	Achieving Agile Big Data Science: The Evolution of a Team's Agile Process Methodology	Ivan Shamshurin
<b>3:35</b>	<b>Coffee Break</b>	
3:55	A Hybrid Approach to Dynamic Enterprise Data Platform	Ahmet Tugrul Bayrak
4:20	A Comparison and an Evaluation between ETL and Data Wrangling Open Source Tools: Exploring Self-Data Preparation Opportunities	Umar Aftab
4:45	Effectiveness of volumetric dataset reduction in testing big data applications	Jaganmohan Chandrasekaran
<b>5:10</b>	<b>Closing Remarks</b>	

4th Workshop on Real-time and Stream Analytics in Big Data & Stream Data Management		
Time	Title	Presenter/Author
9:00 - 9:10	Workshop opening - Introduction	Sabri SKHIRI - EURA NOVA
9:10- 9:50	Keynote 1: Apache Pulsar - Pub-Sub, Storage and Compute with FaaS	Mateo Merli - Steamlio (Splunk)
9:50- 10:30	Keynote 2: Kafka Stream & Evolution of Streaming paradigms.	John Roesler - Confluent
10:30 - 11:00	<b>Coffee Break</b>	
11:00 - 11:25	Scalable and Reliable Multi-Dimensional Aggregation of Sensor Data Streams	Sören Henning and Wilhelm Hasselbring,
11:25 - 11:50	Performance Characterization and Modeling of Serverless and HPC Streaming Applications	Andre Luckow and Shantenu Jha
12:15 - 14:00	<b>Lunch Break</b>	
14:00 - 14:25	Collaborative Streaming: Trust Requirements for Price Sharing	Tobias Grubenmann, Daniele Dell'Aglio, and Abraham Bernstein,
14:50 - 15:15	Kennard-Stone Balance Algorithm for Time-series Big Data Stream Mining	Tengyue Li, Simon Fong, and Raymond Wong
15:15 - 15:45	Assessing the Effects of TV Ad Events on Digital Search: On the Selection of Outcome Measures	Shawndra Hill, Anthony Colas, H. Andrew Schwartz, and Gordon Burtch
15:45 - 16:10	<b>Coffee Break</b>	
16:10 - 16:35	MLK Smart Corridor: An Urban Testbed for Smart City Applications"	Austin Harris, Jose Stovall, and Mina Sartipi
16:35 - 17:00	Image Mining for Real Time Quality Assurance in Rapid Prototyping	Sebastian Trinks and Carsten Felde
17:00 - 17:25	Real-Time Machine Learning Competition on Data Streams at the IEEE Big Data 2019	Dihia Boulegane
	<b>Closing Remarks</b>	

Applications of Big Data Technology in the Transport Industry <i>Workshop Chair: John Easton – University of Birmingham, UK</i>		
Time	Title	Presenter/Author
13:30	Travel Pattern Extraction from Smart Card Data using Data Polishing	Mio Hosoe, Masashi Kuwano, Taku Moriyama, Kosuke Miyazaki and Masaki Ito
13:55	Classifying In-vehicle Noise from Multi-channel Sound Spectrum by Deep Beamforming Networks	Seok-Jun Bu and Sung-Bae Cho
14:20	Analysis of Safety of the Intended Use (SOFIT)	Edward Schwalb
14:45	A Deep Learning Approach to Trespassing Detection Using Video Surveillance Data	Muzammil Bashir, Elke Rundensteiner and Ramoza Ahsan
15:10	Blockchain Technology as a Mechanism for Digital Railway Ticketing	Joseph Preece and John Easton
<b>15:35</b>	<b>Coffee Break</b>	
15:55	Agent-based Modelling to Simulate Road Travel Using Big Data from Smartphone GPS: An Application to the Continental United States	Sashikanth Gurram, Vijayaraghavan Sivaraman, Jonathan Apple and Abdul Pinjari
16:20	Modelling Dynamic Spatial-temporal Cluster Relationships	Ivens Portugal, Paulo Alencar and Cowan Donald
16:45	Using Timed Sequential Patterns in the Transportation Industry	Somayah Karsoum, Le Gruenwald, Clark Barrus and Eleazar Leal

17:10	Analysis of Hazards for Autonomous Driving	Edward Schwalb
17:35	Personalized POI Embedding for Successive POI Recommendation with Large-scale Smart Card Data	Jin-Young Kim, Kyung-Hyun Lim and Sung-Bae Cho
18:00	Closing Remarks	

The Third IEEE International Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD), Wednesday, December 11		
Workshop Chairs: Zhiyuan Chen, Jianwu Wang, Feng Chen, Yiming Ying		
Workshop website: <a href="https://userpages.umbc.edu/~jianwu/BPOD">https://userpages.umbc.edu/~jianwu/BPOD</a>		
Time	Title	Presenter/Authors
	Session 1: Benchmark	
10:05-10:20	An Empirical Study of Rabin Fingerprinting Parameters	Paul Lu, Owen Randall, and Emma McDonald
10:20-10:40	Benchmarking the Discretisation Level of Continuous Attributes: Theoretical and Experimental Approaches	Wanghu Chen, Chao Wang, Jing Li, Bo Yang, Yang Liu, and Jianwu Wang
10:40-11:00	Performance Benchmarking of Data Argumentation and Deep Learning for Tornado Prediction	Carlos Barajas, Matthias Gobbert, and Jianwu Wang
11:00-11:20	Mix and Rank: A Framework for Benchmarking Recommender Systems	Bibek Paudel, Dragi Koccev, and Tome Eftimov
11:20-11:40	GeoYCSB: A Benchmark Framework for the Performance and Scalability Evaluation of NoSQL Databases for Geospatial Workloads	Suneuy Kim and Yuvraj Kanwar
11:40-11:55	Towards a High-Level Description for Generating Stream Processing Benchmark Applications	Alessio Pagliari, Fabrice Huet, and Guillaume Urvoy-Keller
11:55-12:10	Measuring, Quantifying, and Predicting the Cost-Accuracy Tradeoff	Matt Baughman, Nifesh Chakubaji, Krists Kreics, Hong-Linh Truong, and Kyle Chard
12:10-2:30	Lunch and Main Conference Keynote	
	Session 2: Analytics	
2:30-2:50	An Empirical Study of Quad-Level Cell (QLC) NAND Flash SSDs for Big Data Applications	Shuwen Liang, Zhi Qiao, Sihai Tang, Song Fu, and Weisong Shi
2:50-3:10	An Experimental Comparison of GPU Techniques for DBSCAN Clustering	Hamza Mustafa, Eleazar Leal, and Le Gruenwald
3:10-3:30	A Gray-box Testing Method for Divide & Conquer in Image Processing	Marco Strutz, Hermann Hessling, and Achim Streit
3:30-3:50	GHOSTZ PW/GF: distributed parallel homology search system for large-scale metagenomic analysis	Kenta Machida and Osamu Tatebe
3:50-4:10	Search for K: Assessing Five Topic Modeling Approaches to 120,000 Canadian Articles	Qiang Fu, Yufan Zhuang, Jiaxin Gu, Yushu Zhu, Huihui Qin, and Xin Guo
4:10-4:30	Coffee Break	
	Session 3: Tuning	
4:30-4:50	Analysis and Prediction of Data Transfer Throughput for Data-Intensive Workloads	Devarshi Ghoshal, Kesheng Wu, Eric Pouyoul, and Erich Strohmaier
4:50-5:10	Fast Stochastic Block Partitioning using a Single Commodity Machine	Md Abdul Motaleb Faysal and Shaikh Arifuzzaman
5:10-5:30	Cluster-size optimization within a cloud-based ETL framework for Big Data	Eftim Zdravevski, Petre Lameski, Ace Dimitrievski, Marek Grzegorowski, and Cas Apanowicz
5:30-5:50	GraphOpt: a Framework for Automatic Parameters Tuning of Graph Processing Frameworks	Muaz Twaty and Amine Ghrab

5:50-6:10	Multidimensional Preference Query Optimization on Infrastructure Monitoring Systems	Yinghua Qin and Gheorghii Guzun
6:10-6:25	Reactive Microservices in Commodity Resources	Divya Goel and Amaresh Nayak

<b>Advances in High Dimensional Big Data</b> <i>Workshop Chairs: Sotiris Tasoulis</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
	Efficient feature embedding of 3D brain MRI images for content-based image retrieval with deep metric learning	Yuto Onga, Shingo Fujiyama, Hayato Arai, Yusuke Chayama, Hitoshi Iyatomi, and Kenichi Oishi

<b>International Workshop on Big data for Intelligent Transportation Systems</b> <i>Workshop Chairs: Aibek Musaev, Steven Jones</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
12:15	<b>Lunch (provided by the conference)</b>	
1:30	<b>Opening Remarks:</b> Workshop format, expected outcomes	Aibek Musaev (UA)
	<b>Presentations</b>	
1:35	Classification of Driving Behavior Events Utilizing Kinematic Classification and Machine Learning for Down Sampled Time Series Data	Vikram Krishnamurthy (Nissan USA)
1:45	Detecting Pedestrian Crossing Events in Large Video Data from Traffic Monitoring Cameras	Weijia Xu (UT Austin)
1:55	Scalable Object Tracking in Smart Cities	Jose Stovall (UT-Chattanooga)
2:05	Training-free Monocular 3D Event Detection System for Traffic Surveillance	Lijun Yu (CMU)
2:15	Preliminary Research on Vehicle Speed Detection using Traffic Cameras	Aibek Musaev (UA)
2:25	Piecewise Stationary Modeling of Random Processes Over Graphs With an Application to Traffic Prediction	Arman Hasanzadeh (TAMU)
2:35	Knowledge and Situation-Aware Vehicle Traffic Forecasting	Hao Peng (UGA)
2:45	Discovering High Demanding Bus Routes Using Farecard Data	Arbee L.P. Chen (Asia University)
2:55	Towards Building an Interactive Platform for Analyzing Movement of Buses in Hajj	Atif Naseer (Umm Al-Qura University)
3:05	Using Governance and Adaptive Normative Multiagent Systems for Dynamic Vehicle Platoon Formation	Lauro Caetano (Pontifical Catholic University)
3:15	802.11ac and p in a Simulated VANET Environment	Michael Lee (UA)
3:25	Exploring Competition Patterns in Automotive Market with Massive Sales Leads	Sheng Zhang (Beijing Normal University)
3:35	<b>Coffee Break</b>	
4:00	Collaboration Work 1: Discuss presented papers.	
4:30	<b>Break</b>	
4:45	Collaboration Work 2: Design draft report on BITS 2019. <ul style="list-style-type: none"> <li>Group 1: Motivation (research challenges)               <ul style="list-style-type: none"> <li>Group 2: Proposed ITS</li> </ul> </li> <li>Group 3: Summary and conclusion</li> </ul>	
5:30	<b>Closing Remarks</b>	

<b>8th Workshop on Scalable Cloud Data Management</b> <i>Workshop Chairs: Felix Gessert, Wolfram Wingerath, Norbert Ritter</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
<b>8:00- 8:10 am</b>	<b>Opening Remarks</b>	Norbert Ritter (University of Hamburg, Germany)
8:10 – 9:10 am	<b>Keynote</b> Rethinking Cloud scheduling and Management in the era of AI	Liqiang Wang (University of Central Florida, US)



9:10 – 10:40 am	Scaling Data Stores with Skewed Data Access: Solutions and Opportunities	Shahram Ghandeharizadeh (University of Southern California, US)
	Hybrid Scalable Action Rule: Rule Based and Object Based	Jaishree Ranganathan (University of North Carolina at Charlotte, US)
	Kafka: the Database Inverted, but Not Garbled or Compromised	Sean Rooney (IBM Research, Switzerland)
<b>10:40 – 11:00 am</b>	<b>Coffee Break</b>	
11:00am – 13:00pm	On the RESTful Web Services for Managing Application Virtualization Environments	Elif Cansu Yildiz (Link Bilgisayar, Turkey)
	A Technical Perspective of DataCalc - Ad-hoc Analyses on Heterogeneous Data Sources	Johannes Luong (Technische Universität Dresden, Germany)
	Energy Efficient Decentralized Geographical Load Balancing via Dynamic Deferral of Workload	<i>Zeenat Islam</i> (Bangladesh University of Engineering and Technology, Bangladesh)
	SEDM: SSD-Enhanced in-Memory Data Management for Big Data	Wenmei Wu (Renmin University of China, China)
<b>13:00 – 13:15 pm</b>	<b>Closing Remarks</b>	

<b>5th IEEE Workshop on Big Data Analytics in Supply Chains and Transportation</b> Chairs: Dr Allan Zhang and Prof Satish Ukkusuri		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
2:00 pm - 2:05 pm	<b>Opening remarks</b>	
2:05 pm - 2:25 pm	<i>A Mobility-Driven Approach to Modeling Building Energy</i>	<i>Anne Berres., et al.</i>
2:25 pm - 2:45 pm	<i>Performance Evaluation of Ethereum-based On-chain Sensor Data Management Platform for Industrial IoT</i>	Kentaroh Toyoda, et al.
2:45 pm – 3:05 pm	<i>Estimation of Transaction Network Data Between Branch Offices using Transaction Big Data Throughout Japan</i>	Yoshiki Ogawa, et al.
3:05 pm – 3:25 pm	<i>Analysis of Hurricane Matthew 2016 Data to Estimate Airlines Passengers Disruption</i>	<i>Harshitha Meda, et al.</i>
3:25pm - 3:45pm	<i>Fast Top-N Personalized Recommendation on Item Graph</i>	<i>Zhuoyi Lin</i>
3:45pm- 4:05pm	<i>Service Time Prediction for Last-Yard Delivery</i>	<i>Joel Wei En Tay</i>
<b>4:05 pm - 4:25 pm</b>	<b>Coffee Break</b>	
4:25 pm - 4:45 pm	<i>Modeling and Analysis of Individual Travel Behavior using License Plate Recognition Data</i>	<i>Yuhuan Lu</i>
4:45 pm - 6:00pm	<b>Poster and Network</b>	
	<i>Coping with Big Data in Transfer Optimization</i>	<i>Mojtaba Shakeri , et al.</i>
	<i>The Blessing of Dimensionality in Many-Objective Search: An Inverse Machine Learning Insight</i>	<i>Abhishek Gupta, et al.</i>
	<i>Ensemble kriging for environmental spatial processes</i>	<i>Gokhan Yagli, et al.</i>
	<b>Closing remarks</b>	

<b>Analysis of Large-scale Disparate Data</b> Workshop Chairs: Dr. Michael Barton, Prof. Dhableswar Panda, Brian Panneton, Dr. Simon Su, Dr. Rhonda Vickery, Dr. Venkat Dasari, Brendan Tauras, Prof. Chen Li, Prof. Bo Sun
--



Time	Title	Presenter/Author
8:00	Opening Remarks	
8:10	Dynamic Collaborative Visualization Ecosystem to Support the Analysis of Large-Scale Disparate Data	Christopher Koehler
8:35	Hybrid 2D and 3D Visual Analytics of Network Simulation Data	Simon Su
9:00	Visualization Techniques for Large-Scale Monte Carlo Simulation	Vincent Perry
9:25	Analysis of High fidelity ns-3 simulations to study real-time application performance in tactical wireless networks	Venkat Dasari
<b>9:50</b>	<b>Coffee Break</b>	
10:10	Dynamic visualization of large scale tactical network simulations	Venkat Dasari
10:35	Detecting Network Soft-failures with the Network Link Outlier Factor (NLOF)	Christopher Mendoza
11:00	Leveraging Comprehensive Data Analysis to Inform Parallel HPC Workloads	Matthew Dwyer
11:25	Information geometry for big and disparate data	Vinod Mishra
<b>11:50</b>	<b>Closing Remarks</b>	

Policy-based Autonomic Data Governance (PADG) Workshop Chairs: Elisa Bertino, Seraphin Calo, Dinesh Verma		
Time	Title	Presenter/Author
08:30 – 08:40	Opening Remarks	
	Keynote	
08:40 – 09:25		TBD
	Session I (Elisa Bertino)	
09:25 – 09:45	Policy based Ensembles for applying ML on Big Data	Dinesh Verma
09:45 – 10:05	A Policy-based Approach for Measuring Data Quality	Seraphin Calo
10:05 – 10:25	Enabling Privacy Policies for mHealth Studies	Brian Wang
10:25 – 10:45	Coffee Break	
	Session II (Seraphin Calo)	
10:45 – 11:05	Access Control Model Extensions to Support Data Privacy Protection based on GDPR	Maryam Davari
11:05 – 11:25	On the Quality of Classification Models for Inferring ABAC Policies from Access Logs	Stefano Valtolina
11:25 – 11:45	Counting Devices: Revisiting Existing Approaches in Today's Settings	Franck Le
11:45 – 12:05	Towards a Neural-Symbolic Generative Policy Model	Daniel Cunningham
12:05 – 12:15	Closing Remarks	
12:15 – 01:30	Lunch	

#### The 4th IEEE International Workshop on Big Spatial Data (BSD 2019)

Start	End	Event
7:00	8:15	Registration
08:15	08:30	Opening and Welcome
08:30	10:15	Session: Data Management
8:30	8:50	<i>Large Scale Spatial Data Processing With User Defined Filters In BboxDB</i> Jan Kristof Nidzwetzki and Ralf Hartmut Güting
8:50	9:10	<i>Euler++: Improved Selectivity Estimation for Rectangular Spatial Records</i> A.B. Siddique, Ahmed Eldawy, and Vagelis Hristidis
9:10	9:30	<i>bench4gis: Benchmarking Privacy-aware Geocoding with Open Big Data</i> Daniel Haris and Chris Delcher
9:30	9:50	<i>Towards Building Evacuation Planning Platform using Multimodal Transportation for a Large Crowd: Work in progress</i> Emad Felemban, Faizan Ur Rehman, Hassan Wadood, and Atif Naseer
9:50	10:15	<i>SEXTANT: A Computational Framework for Scalable and Efficient Correlation of Spatio-Temporal Trajectories</i> Brian Thompson, Dave Cedel, Jeremy Martin, Kristen Snee, and Alex Cheung
10:15	10:30	Coffee Break (California Foyer)
10:30	11:50	Session: Pattern Discovery
10:30	10:50	<i>Accurate Spatial Mapping of Social Media Data with Physical Locations</i> Mohit Mittal, Panote Siriaraya, Chonho Lee, Yukiko Kawai, Takashi Yoshikawa, and Shinji Shimojo
10:50	11:10	<i>Co-location Pattern Mining of Geosocial Data to Characterize Urban Functional Spaces</i> Arif Masrur, Gautam Thakur, Kevin Sparks, Rachel Palumbo, and Donna Peuquet
11:10	11:35	Monitoring of natural disasters through anomaly detection on mobile phone data

		Aude Marzuoli and Fengmei Liu
11:35	11:55	<i>User Identification across Online Social Networks Based on Similarities among Distributions of Friends' Locations</i> Kazufuhiro Tanimi Kojima, Keisuke Ikeda, and Masahiro Tani
12:00	1:30	Lunch (On Your Own)
1:30	3:40	Session: Learning and Data Mining
1:30	1:50	<i>Evaluation of Location Estimation Method That Focuses on Geographical Proximity of Friends</i> Keisuke Ikeda, Kazufumi Kojima, and Masahiro Tani
1:50	2:15	<i>Mining top-up transactions and online classified ads to predict urban neighborhoods socioeconomic status</i> Eduardo Cruz, Carmen Vaca, and Allan Avendaño
2:15	2:35	<i>Spatio-temporal classification at multiple resolutions using multi-view regularization</i> Guruprasad Nayak, Rahul Ghosh, Vipin Kumar, Xiaowei Jia, and Varun Mithal
2:35	3:40	Keynote: <a href="#">Li Xion</a> (Emory University)
3:40	4:00	Coffee Break (California Foyer)
4:00	5:40	Session: Data Analysis and Visualization
4:00	4:25	<i>Decision-Making System for Road-Recovery Considering Human Mobility by Applying Deep Q-Network</i> SooHyun JOO, Yoshiki Ogawa, and Yoshihide Sekimoto
4:25	4:45	<i>cPSITRES: A collaborative system for analysis of Big Data on sea ice</i> Vinit Veerenraveer Singh, Scott Sorensen, and Chandra Kambhamettu
4:45	5:05	<i>A demonstration of B-EagleV: Visualizing massive point cloud directly from HDFS</i> Minh Hieu Nguyen, Sanghyun Yoon, Sangyoon Park, and Joon Heo
5:05	5:30	<i>BigDataCube: A Scalable, Federated Service Platform for Copernicus</i> Dimitar Mišev, Peter Baumann, Dimitris Bellos, and Wiehrend Stef
5:30	6:00	Adjourn and Closing

**Second International Workshop on the Internet of Things Data Analytics (IoTDA)***Tuesday – December 10, 2019 - Location: TBD**Workshop Chairs: Eyhab Al-Masri and Yan Bai (University of Washington Tacoma)*

Time	Title	Presenter/Author
<b>11:00 am – 11:05 am</b>	<b>Opening Remarks and Welcome (Eyhab Al-Masri)</b>	
11:10 am– 11:30 am	Ptolemaic Indexing for Managing and Querying Internet of Things (IoT) Data	Christian Beecks, Fabian Berns, and Kjeld Schmidt
11:30 am– 11:50 am	Smart Home IoT Anomaly Detection based on Ensemble Model Learning From Heterogeneous Data	Sihai Tang, Zhaochen Gu, Song Fu, and Qing Yang
11:50 am– 12:10 pm	Automatic Hyperparameter Tuning Method for Local Outlier Factor, with Applications to Anomaly Detection	Zekun Xu, Deovrat Kakde, and Arin Chaudhuri
<b>12:10 pm – 1:30 pm</b>	<b>Lunch</b>	
1:40 pm– 2:00 pm	Kensor: Coordinated Intelligence from Co-Located Sensors	Olivera Kotevska, Kalyan Perumalla, and Juan Lopez Jr.
2:00 pm– 2:20 pm	A New Approach for Efficient Structure Discovery in IoT	Fabian Berns, Kjeld Schmidt, Alexander Grass, and Christian Beecks
2:20 pm– 2:40 pm	Cost-adaptive Neural Networks for Peak Volume Prediction with EMM Filtering	Bin Yu, Giovanna Graciani, Anderson Nascimento, and Juhua Hu
<b>2:40 pm – 2:55 pm</b>	<b>Coffee Break</b>	
3:00 pm– 3:20 pm	Speech Emotion Detection using IoT based Deep Learning for Health Care	Zeenat Tariq, Sayed Khushal Shah, and Yugyung Lee
3:20 pm– 3:40 pm	Effectively Testing System Configurations of Critical IoT Analytics Pipelines	Morgan Geldenhuys, Lauritz Thamsen, Kain Kordian Gontarska, Felix Lorenz, and Odej Kao
3:40 pm– 4:00 pm	Detecting errors in short-term electricity demand forecast using people dynamics	Guillaume Habault, Yasutaka Nishimura, Kiyohito Yoshihara, and Chihiro Ono
<b>4:00 pm – 4:20 pm</b>	<b>Coffee Break</b>	
4:20 pm– 4:40 pm	Towards an Interactive Visualization Framework for IoT Device Data Flow	Ebelechukwu Nwafor and Habeeb Olufowobi
4:40 pm– 5:00 pm	Association Model between Visual Feature and AQI Rank Using Lifelog Data	Phuong-Binh Vo, Trong-Dat Phan, Minh-Son Dao, and Koji Zettsu
5:00 pm– 5:20 pm	IoT based Urban Noise Monitoring in Deep Learning using Historical Reports	Sayed Khushal Shah, Zeenat Tariq, and Yugyung Lee
<b>5:30 pm – 6:00 pm</b>	<b>Closing Remarks</b>	

**7th International Workshop on Distributed Storage and Blockchain Technologies for Big Data***Workshop Chairs: Hui Li and Kenneth Shum*

Time	Title	Presenter/Author
8:00 – 8:20 am	Predicting Transaction Latency with Deep Learning in proof-of-work Blockchains	Enrico Tedeschi, Tor-Arne S. Nordmo, Dag Johansen and Havard D. Johansen
8:20 – 8:40 am	AONT-NZZD: A Secure and Efficient Dispersal Scheme in Distributed Storage Systems	Pengcheng Xie, Hui Li, Haiyang Yu, and Zequan Chen
8:40 – 9:00 am	Scalability Analysis of Blockchain on a Serverless Cloud	Alex Kaplunovich, Karuna P. Joshi, and Yelena Yesha
	<b>Closing Remarks</b>	

**The 3rd International Workshop on Big Data Analytic for Cyber Crime Investigation and Prevention***Monday, December 9th, 2019; Location: TBD*

Workshop Chairs: Andrii Shalaginov, Jan William Johnsen, Ambika Shrestha Chitrakar and Asif Iqbal Norwegian University of Science and Technology, Gjovik, Norway and KTH Royal Institute of Technology in Stockholm, Sweden		
Time	Title	Presenter/Authors
08:00-08:05	<b>Opening Remarks and Welcome</b>	Andrii Shalaginov
08:05-08:20	<i>Cybercrime Investigations in the Era of Smart Applications: Way Forward Through Big Data</i>	Andrii Shalaginov, Igor Kotsiuba, Asif Iqbal
08:20-09:00	<b>Keynote speech</b> <i>Forensic Readiness as an inevitable part of open, cooperative and smart cyber-physical systems</i>	Igor Kotsiuba
09:00-09:20	<i>Preprocessing for open-source intelligence gathering in criminal investigations</i>	Jan William Johnsen and Katrin Franke
09:20-09:40	<i>Comparative Study of Wear-leveling in Solid-State Drive with NTFS File System</i>	Ashar Neyaz and Bing Zhou
09:40-10:00	<i>Technique for Finding and Investigating the Strongest Combinations of Cyberattacks on Smart Grid Infrastructure</i>	Igor Kotsiuba, Inna Skarga-Bandurova, Alkiviadis Giannakoulis, Mykhailo Chaikin, and Aleksandar Jevremovic
<b>10:00-10:30</b>	<b>Coffee Break</b>	
10:30-10:50	<i>PACE: Platform for Android Malware Classification and Performance Evaluation</i>	Ajit Kumar, Vinti Agarwal, Shishir K. Shandilya, Andrii Shalaginov, Saket Upadhyay, and Bhavna Yadav
10:50-11:10	<i>Detecting Web Spams in Webgraphs with Predictive Model Analysis</i>	Naw Safrin Sattar, Shaikh Arifuzzaman, Minhaz F. Zibran, and Md Mohiuddin Sakib
11:10-11:30	<i>Detection of Fraudulent Behavior Using the Combined Algebraic and Machine Learning Approach</i>	Oleksandr Letychevskyi and Tetiana Polhul
11:30-11:50	<i>Security validation testing environment in the cloud</i>	Cristian Bucur and Eduard Babulak
11:50-12:10	<i>Deep in the Dark: A Novel Threat Detection System using Darknet Traffic</i>	Sanjay Kumar, Harald Vranken, Joost Dijk, and Timo Hamalainen
12:10-12:30	<i>Basic Forensic Procedures for Cyber Crime Investigation in Smart Grid Networks</i>	Igor Kotsiuba, Oksana Bulda, Inna Skarga-Bandurova, and Alkiviadis Giannakoulis
<b>12:30-12:40</b>	<b>Closing Remarks</b>	Andrii Shalaginov

The Third Annual Workshop on Applications of Artificial Intelligence in the Legal Industry Workshop Chairs: Jianping Zhang, Nathaniel Huber-Fliflet, Robert Keeling, Christian J. Mahoney, Haozhen Zhao		
Time	Title	Authors
1:00-1:30pm	<b>Opening Session</b>	
1:30-2:00pm	Japanese Mistakable Legal Term Correction using Infrequency-aware BERT Classifier	Takahiro Yamakoshi, Takahiro Komamizu, Yasuhiro Ogawa, et al.
2:00-2:30pm	Normalisation of SWIFT Message Counterparties with Feature Extraction and Clustering	Thanasis Schoinas, Benjamin Guinard, Diba Esbati, et al.
2:30-3:00pm	Experimental Evaluation of CNN Parameters for Text Categorization in Legal Document Review	Qian Han, Yufeng Kou, and Derek Snidauf
3:00-3:30pm	Forecasting of Trends in Legal Spend Management	Pragati Awasthi, Jerzy Bala, and Sebastian Carter
3:30-4:00pm	<b>Coffee Break</b>	
4:00-4:30pm	Image Analytics for Legal Document Review: A Transfer Learning Approach	Nathaniel Huber-Fliflet, Fusheng Wei, Haozhen Zhao, et al.
4:30-5:00pm	Qualitative Mapping Modeling of Criminals' Sense of Security in Theft Cases	Yajie Su, Hao Lu, and Hongzhou Zhang
5:00-5:30pm	Supervised Key Terms Clustering for Regulatory Monitoring	Yong Zou and Andrew (Pat) Waldo
5:30-6:00pm	<b>Closing Session</b>	

Big Data Predictive Maintenance using Artificial Intelligence Workshop Chairs: Ryan Benton, Rituparna Data, and Aviv Segev
---

<b>Date: 09 December 2019</b>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
8:00-8:05	Opening Remarks	
8:05-8:30	Machine Learning Use Cases for Smart Manufacturing KPIs	Sandeep Jeeretty, Ken Kennedy, Edward Duffy, Annie Walker, and Bennie Vorster
8:30-8:55	Forecasting cross-border power exchanges through an HVDC line using dynamic modelling	Sylvie Koziel, Patrik Hilber, Per Westerlund, and Ebrahim Shayesteh
8:55-9:20	Failing & !Falling (F&!F): Learning to Classify Accidents and Incidents in Aircraft Data	Jarrold Carson, Kane Hollingsworth, Rituparna Datta, and Aviv Segev
9:20-9:55	Data Imputation Method based on Programming by Example: APREP-S	Hiroko Nagashima and Yuka Kato
9:55-10:20	Application of Machine Learning and Spatial Bootstrapping to Image Processing for Predictive Maintenance	Vikram Krishnamurthy, Kusha Nezafati, and Vikrant Singh
10:20-10:40	<b>Coffee Break</b>	
10:40-11:05	Self-supervised Multi-stage Estimation of Remaining Useful Life for Electric Drive Units	Ivan Melendez, Rolando Dölling, and Oliver Bringmann
11:05-11:30	Wind Turbine operational state prediction: towards featureless, end-to-end predictive maintenance	Adrian Stetco, Anees Mohammed, Siniša Djurović, Goran Nenadic, and John Keane
11:30-11:55	Prescriptive Equipment Maintenance: A Framework	Suresh Choubey, Ryan Benton, and Tom Johnsten
11:55-12:20	Spatiotemporal Real-Time Anomaly Detection for Supercomputing Systems	Qiao Kang, Ankit Agrawal, Alok Choudhary, Alex Sim, Kesheng Wu, Rajkumar Kettimuthu, Peter Beckman, Zhengchun Liu, and Wei-keng Liao
<b>12:20-12:30</b>	<b>Closing Remarks</b>	

<b>The 2nd International Workshop on Big Data for Marketing Intelligence and Operation Management</b>		
<i>Workshop Chairs: Wutao Wei, Huaiye Zhang, Yiwen Zhang</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
10:30-11:00	Purchase Prediction for Paying Players in Free Online Games via Survival Analysis	Wanshan Yang, Ting Huang, Junlin Zeng, Yan Tang, Lijun Chen, Shivakant Mishra, and Youjian (Eugene) Liu
11:00-11:30	Subject-Oriented Data Retrieval and Analysis on Sina Weibo	Dan Lo
11:30-12:00	Seasonality-Adjusted Conceptual-Relevancy-Aware Recommender System in Online Groceries	Luyi Ma, Jason H.D. Cho, Sushant Kumar, and Kannan Achan
	<b>Closing Remarks</b>	

<b>2nd Workshop on Energy-Efficient Machine Learning and Big Data Analytics</b>		
<i>Workshop Chair: Mohammed Alawad</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
9:00-9:25 am	Evolving Energy Efficient Convolutional Neural Networks	Steven Young, Travis Johnston, Catherine Schuman, Pravallika Devineni, Bill Kay, Derek Rose, Maryam Parsa, Robert Patton, and Thomas Potok
9:25-9:50 am	An Energy-Efficient Reconfigurable LSTM Accelerator for Natural Language Processing	Elham Azari and Sarma Vrudhula
9:50-10:10 am	Exploration of OpenCL 2D Convolution Kernels on Intel FPGA, CPU, and GPU Platforms	Zheming Jin and Hal Finkel
10:10-10:30 am	<b>Coffee Break</b>	
10:30-10:55 am	Bayesian-based Hyperparameter Optimization for Spiking Neuromorphic Systems	Maryam Parsa, J Parker Mitchell, Catherine Schuman, Robert Patton, Thomas Potok, and Kaushik Roy
10:55-11:20 am	A Case Study of k-means Clustering using SYCL	Zheming Jin and Hal Finkel
	<b>Closing Remarks</b>	

**The First International Workshop on Big Data Tools, Methods, and Use Cases for Innovative Scientific Discovery (BTSD'19)**

Workshop Chairs: Matt Sangkeun Lee, Travis Johnston DATE: December 9		
Time	Title	Presenter/Author
1:30pm	Welcome: Where are we at? What is missing?	Matt Lee/Travis Johnston
1:50pm	Identifying Time Series Similarity in Large-Scale Earth System Datasets	Payton Linton, William Melodia, Alina Lazar, Deborah Agarwal, Ludovico Bianchi, Devarshi Ghoshal, Kesheng Wu, Gilberto Pastorello, and Lavanya Ramakrishnan
2:10pm	Learning to Predict Material Structure from Neutron Scattering Data	Cristina Garcia Cardona, Ramakrishnan Kannan, Travis Johnston, Thomas Proffen, Katharine Page, and Sudip Seal
2:30pm	Realistic Transport Simulation: Tackling the Small Data Challenge with Open Data	Guimu Guo, Jalal Majed Khalil, Da Yan, and Virginia Sisiopiku
2:50pm	Information Extraction from Cancer Pathology Reports with Graph Convolution Networks for Natural Language Texts	Hong-Jun Yoon, John Gounley, M. Todd Young, and Georgia Tourassi
3:10pm	Machine Learning for Prediction of Mid to Long Term Habitual Transportation Mode Use	Alina Lazar, Alexandra Ballow, Ling Jin, C. Anna Spurlock, Alex Sim, and Kesheng Wu
3:40pm	<b>Coffee Break</b>	
4:00pm	Clustered Latent Dirichlet Allocation for Scientific Discovery	Christopher Gropp, Alexander Herzog, Ilya Safro, Paul Wilson, and Amy Apon
4:20pm	Quantum Grover search-based optimization for innovative material discovery	Sima Esfandiarpour Borujeni, Ramkumar Harikrishnakumar, and Saideep Nannapaneni
4:40pm	Detecting Dependency Between Discrete Random Variables and Application	Edgar Llamas, Ivan García, and Andrés Méndez
5:00pm	Visualization System for Evolutionary Neural Networks for Deep Learning	Junghoon Chae, Catherine Schuman, Steven Young, Travis Johnston, Derek Rose, Robert Patton, and Thomas Potok
5:20pm	<b>Session Break</b>	
5:40pm	Exploration of Workflow Management Systems Emerging Features from Users Perspectives	Ryan Mitchell, Loic Pottier, Steve Jacobs, Rafael Ferreira da Silva, Mats Rynge, Karan Vahi, and Ewa Deelman
6:00pm	Empowering Agroecosystem Modeling with HTC Scientific Workflows: The Cycles Model Use Case	Rafael Ferreira da Silva, Rajiv Mayani, Yuning Shi, Armen R. Kemanian, Mats Rynge, and Ewa Deelman
6:20pm	Evaluating Scientific Workflow Engines for Data and Compute Intensive Discoveries	Rina Singh, Jeffrey Graves, Sreenivas Sukumar, and Valentine Anantharaj
6:40pm	Wrap up: What's next?	Matt Lee/Travis Johnston
	<b>Closing Remarks</b>	

<b>The Third IEEE Workshop on Human-in-the-loop Methods and Human-Machine Collaboration in BigData (HMDData 2019)</b> <i>Workshop Chair: Senjuti Basu Roy (New Jersey Institute of Technology), Alex Quinn (Purdue University), Atsuyuki Morishima (University of Tsukuba)</i> <i>The latest program is available at <a href="https://humanmachinedata.org">https://humanmachinedata.org</a></i>		
Time	Title	Author
8:50	<b>Opening</b>	
8:55	Human-Machine Information Extraction Simulator for Biological Collections	Icaro Alzuru, Aditi Malladi, Andréa Matsunaga, Maurício Tsugawa, and José A.B. Fortes
9:20	Establishment of Work-Flow for Roof Damage Detection Utilizing Drones, Human and AI based on Human-in-the-Loop Framework	Munenari Inoguchi, Keiko Tamura, and Ryota Hamamoto
9:45	Misinformation Harms During Crises : When The Human and Machine Loops Interact	Thi Tran, Rohit Valecha, Paul Rad, and H.Raghav Rao
9:57	Analysing Social Media as a Hybrid Tool to Detect and Interpret likely Radical Behavioural Traits for National Security	Pedro Cárdenas, Georgios Theodoropoulos, Boguslaw Obara, and Ibad Kureshi
10:10	<b>Coffee Break</b>	
10:30	Collaborative Workflow for Analyzing Large-Scale Data for Antimicrobial Resistance: An Experience Report	Pei-Yu Hou, Jing Ao, Andrew Rindos, Shivaramu Keelara, Paula J. Fedorka-Cray, and Rada Chirkova
10:55	Incentive Design for Crowdsourced Development of Selective AI for Human and Machine Data Processing: A Case Study	Masafumi Hayashi, Masaki Kobayashi, Masaki Matsubara, Toshiyuki Amagasa, and Atsuyuki Morishima



11:20	Active Learning Strategies for Hierarchical Labeling Microtasks	Kousuke Uo, Masaki Kobayashi, Masaki Matsubara, Yukino Baba, and Atsuyuki Morishima
11:45	Disambiguation and Error Resolution in Call Transcripts	Jordan Hosier, Vijay K.Gurbani, and Neil Milsted
12:10	A Microtask Approach to Identifying Incomprehension for Facilitating Peer Learning	Hinako Izumi, Masaki Matsubara, Chiemi Watanabe, and Atsuyuki Morishima
12:25	<b>Lunch</b>	
14:00	<b>Keynote:</b> Computation and Organization	Michael Bernstein (Stanford University)
15:00	<b>Short Break</b>	
15:10	ClusterClean: a Weak Semi- Supervised Approach for Cleaning Data Labels	Kyriaki Dimitradou, Rahul Manghwani, and Timothy C.Hoad
15:25	Towards Quality Assessment of Crowdsourcing Output Based on Behavioral Data	Shigeaki Yuasa, Takumi Nakai, Takanori Maruichi, Manuel Landsmann, Koichi Kise, Masaki Matsubara, and Atsuyuki Morishima
15:40	<b>Coffee Break</b>	
16:00	Extracting Explainable Deep Representation for Machine Tutoring	Ming-Chen Wang, Vahid Golderzahi, and Hsing-Kuo Pao
16:25	Explainable Recommendation Using Review Text and a Knowledge Graph	Takafumi Suzuki, Satoshi Oyama, and Masahito Kurihara
16:37	WATAPI: Composing Web API Specification from API Documentations through an Intelligent and Interactive Annotation Tool	Mehsi Bahrami and Wei-Pend Chen
16:49	Active learning without unlabeled samples: generating questions and labels using Monte Carlo Tree Search	Sathish K. Sankarpandi, Spyros Samothrakis, Luca Citi, and Peter Brady
17:01	Super Long Interval Time-Lapse Image Generation for Proactive Preservation of Cultural Heritage Using Crowdsourcing	Hidehiko Shishido, Hansung Kim, and Itaru Kitahara
17:13	<b>Closing Remarks</b>	

<b>4th Workshop on Open Science in Big Data (OSBD)</b> <i>Workshop Chairs: Shannon Quinn, Michael Cotterell, Kyle Johnsen, John Miller, Suchi Bhandarkar, Nicole Lazar</i> <i>Emcee: Nicholas Klepp</i>		
Time	Title	Presenter/Author
8:00am – 8:05	Welcome	Shannon Quinn
8:05 – 8:50	<i>Scaling Reproducible Research with Jupyter</i>	Carol Willing
8:50 – 9:35	<i>Responsibly Designing Open Science Data Teams</i>	Amanda Casari
9:35 – 10:10	<i>iEnvironment: Perspectives on Metadata-Oriented Testing of Research Software</i>	Doug Mulholland, Paulo Alencar, and Cowan Donald
10:10 – 10:30	<b>Coffee Break</b>	
10:30 – 11:05	<i>Open Source Innovation in Practice: A Lean-Based Development Process Leveraging Open Source Big Data Tools</i>	Silvio Alonso, Marx Viana, Elder Cirilo, Paulo Alencar, and Carlos Lucena
11:05 – 11:40	<i>An Unbiased Benchmark for Bangla Handwritten Digits Recognition</i>	Aminul Islam, Fuad Rahman, and AKM Shahariar Azad Rabby
11:40 – 12:15pm	<i>Using Collaborative Open Science to Advance K-12 Computing Education</i>	Monica McGill
12:15 – 12:50	<i>Code Convention Adherence in Big Data Research Infrastructure: An Exploratory Study</i>	Michael Smit
12:50 – 1:25	<i>Using Synthetic Data Generators to Promote Open Science in Higher Education Learning Analytics</i>	Mohsen Dorodchi, Erfan Al-Hossami, Aileen Benedict, and Elise Demeter
1:25 – 1:30	<b>Closing Remarks</b>	

<b>3rd International Workshop on Big Data Analytics for Cyber Intelligence and Defense (BDA4CID)</b> <i>Workshop Chairs: Stephen McGough</i>		
Time	Title	Presenter/Author
1:30 – 1:40 pm	Arrivals and welcome	Stephen McGough

1:40 – 2:10pm	A Location Independent Machine Learning Approach for Early Fake News Detection	<b>Haohui Liu</b>
2:10 – 2:40pm	IoCMiner: Automatic Extraction of Indicators of Compromise from Twitter	<b>Amirreza Niakanlahiji</b> , Lida Safarnejad, Reginald Harper and Bei-Tseng Chu
2:40 – 3:10 pm	Online Hate A Study on the Feasibility to Detect Hate Speech in Swedish	<b>Johan Fernquist</b> , Oskar Lindholm, Lisa Kaati and Nazar Akrami
3:10 – 3:40pm	PRAT - a Tool for Assessing Risk in Written Communication	<b>Amendra Shrestha</b> , Lisa Kaati and Nazar Akrami
3:40 – 4:00 pm	<b>Coffee Break</b>	
4:00 – 4:30pm	Modeling and Forecasting Armed Conflict: AutoML with Human-Guided Machine Learning	Vito D’Orazio, James Honaker, Raman Prasad, <b>Michael Shoemate</b>
4:30 – 5:00pm	Class Balancing for Fraud Detection in Point of Sale Systems	<b>Christine Hines</b> and Abdou Youssef
5:00 – 5:30pm	Identifying Android Malware Families Using Android-Oriented Metrics	<b>William Blanc</b> , Lina G. Hashem, Karim O. Elish, and Hussain M. J. Almohri
5:30 – 6:00pm	Volenti non fit injuria: Ransomware and its Victims	Amir Atapour-Abarghouei, Stephen Bonner and Andrew Stephen McGough
<b>6:00 pm</b>	<b>Closing Remarks</b>	

<b>Sixth International Workshop on High Performance Big Graph Data Management, Analysis, and Mining</b> <i>Workshop Chairs: Kamesh Madduri, Mohammad Al Hasan, Nesreen Ahmed, and Shaikh Arifuzzaman</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
8:15 – 8:20	Opening Remarks	Shaikh Arifuzzaman and Mohammad Hasan
8:20 – 9:20	Keynote Talk	TBD
9:25 – 9:45	Computing Complex Graph Properties with SQL Queries	Xiantian Zhou and Carlos Ordonez
9:45 – 10:05	Network Embedding: On Compression and Learning	Esra Akbas and Mehmet Aktas
10:05 – 10:25	Distributed Community Detection in Large Networks using an Information-Theoretic Approach	Md Abdul Motaleb Faysal and Shaikh Arifuzzaman
<b>10:25 – 10:45</b>	<b>Coffee Break</b>	
11:00 – 11:20	A Scalable Graph Analytics Framework for Programming with Big Data in R	S. M. Shamimul Hasan, Drew Schmidt, Ramakrishnan Kannan, and Neena Imam
11:20 – 11:40	Efficient similarity-based alignment of temporally-situated graph nodes with Apache Spark	Hubert Naacke, Ke Li, Bernd Amann, and Olivier Curé
11:40 – 12:00	GraphEvo: Characterizing and Understanding Software Evolution using Call Graphs	Vijay Walunj, Gharib Gharibi, Duy Ho, and Yugyung Lee
<b>12:00 – 12:10</b>	<b>Closing Remarks</b>	

<b>The 6th International Workshop on Big Data Analytic Technology for Bioinformatics and Health Informatics (KDDBHI2019)</b> <i>Workshop Chairs: Donghui Wu and Xin Deng</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
1:00 – 1:10	<i>Chairs’ Remarks</i>	Donghui Wu, etc.
1:10 – 1:30	<i>Explainable Deep Learning Applied to Understanding Opioid Use Disorder and Its Risk Factors</i>	<b>Terri Workman</b> , Yijun Shao, Joel Kupersmith, Friedhelm Sandbrink, Joseph Goulet, Nawar Shara, Christopher Spevak, Cynthia Brandt, Marc Blackman, and Qing Zeng-Treitler
1:30 – 1:50	<i>Towards Explainable Melanoma Diagnosis: Prediction of Clinical Indicators Using Semi-supervised and Multi-task Learning</i>	<b>Seiya Murabayashi</b> and Hitoshi Iyatomi
1:50 – 2:10	<i>Computer-Aided Clinical Skin Disease Diagnosis Using CNN and Object Detection Models</i>	Xin He, Shihao Wang, Shaohuai Shi, Zhenheng Tang, Yuxin Wang, Ronghao Ni, Zhihao Zhao, Jing Dai, Xiaofeng Zhang, Xiaoming Liu, Zhili Wu, Wu Yu, and <b>Xiaowen Chu</b>
2:10 – 2:30	<i>Stochastic Gastric Image Augmentation for Cancer Detection from X-ray Images</i>	<b>Hideaki Okamoto</b> , Quan Huu Cap, Takakiyo Nomura, Hitoshi Iyatomi, and Jun Hashimoto



2:30 – 2:50	<i>Automated Machine Learning for EEG-Based Classification of Parkinson's Disease Patients</i>	<b>Milan Koch</b> , Victor Geraedts, Hao Wang, Martijn Tannemaat, and Thomas Bäck
2:50 – 3:10	<i>Recurrent Neural Network Based Feature Selection for High Dimensional and Low Sample Size Micro-array Data</i>	Shanta Chowdhury, <b>Xishuang Dong</b> , and Xiangfang Li
3:10 – 3:30	<i>Exploiting Anti-Monotonic Constraint for Mining Palindromic Motifs from Big Genomic Data</i>	Oluwafemi Sarumi and <b>Carson Leung</b>
3:30 – 3:40	<i>Questions and Discussion</i>	
3:40 – 4:00	<b>Coffee Break</b>	
4:00 – 4:20	<i>Reinforcement Learning Framework to Identify Cause of Diseases - Predicting Asthma Attack Case</i>	<b>Quan Do</b> , Alexa Doig, and Cao Son Tran
4:20 – 4:40	<i>Bayesian Non-linear Support Vector Machine for High-Dimensional Data with Incorporation of Graph Information on Features</i>	<b>Wenli Sun</b> , Changgee Chang, and Qi Long
4:40 – 5:00	<i>Predicting Post-stroke Hospital Discharge Disposition Using Interpretable Machine Learning Approaches</i>	<b>Jin Cho</b> , Alnour Alharin, Zhen Hu, Nancy Fell, and Mina Sartipi
5:00 – 5:20	<i>Discovering Sublanguages in a Large Clinical Corpus through Unsupervised Machine Learning and Information Gain</i>	<b>Terri Workman</b> , Guy Divita, and Qing Zeng-Treitler
5:20 – 5:40	<i>Contrast-resolution Evaluation of Fourier Based High Frame Rate Imaging</i>	<b>Zhaohui Wang</b>
5:40 – 6:00	<i>Questions and Discussion</i>	
	<b>Closing Remarks</b>	

<b>GTA<sup>3</sup> 3.0: The 3rd workshop on Graph Techniques for Adversarial Activity Analytics</b>		
<i>Workshop Chairs: : Jiejun Xu, Hanghang Tong, Andrea Bertozzi, Vince Lyzinski, George Chin, Joel Douglas</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
1:25pm - 1:30pm	Opening Remarks	Workshop Organizers
1:30pm - 2:10pm	Keynote 1	TBD
2:10pm – 2:50pm	Keynote 2	TBD
2:50pm – 3:05pm	Noisy Subgraph Isomorphisms on Multiplex Networks	Hui Jin, Xie He, Yanghui Wang, Hao Li, and Andrea Bertozzi
3:05pm – 3: 20pm	Multiplex Graph Matching Matched Filters	Konstantinos Pantazis, Daniel L. Sussman, Youngser Park, Carey E. Priebe, and Vince Lyzinski
3:20pm – 3:30pm	Graph Generation with a Focusing Lexicon	Mayanka Chandra Shekar and Joseph Cottam
3:40pm – 4:00pm	<b>Coffee Break</b>	
4:00pm – 4:40pm	Keynote 3	TBD
4:40pm – 5:20pm	Keynote 4	TBD
5:20pm – 5:35pm	Applications of Structural Equivalence to Subgraph Isomorphism on Multichannel Multigraphs	Thien Nguyen, Dominic Yang, Yurun Ge, Hao Li, and Andrea Bertozzi
5:35pm – 5:45pm	Filtering Strategies for Inexact Subgraph Matching on Noisy Multiplex Networks	Alexei Kopylov and Jiejun Xu
5:45pm – 5:55pm	Higher Order Temporal Analysis of Global Terrorism Data	Madelyn Dunning and Sumit Purohit
5:55pm – 6:05pm	<b>Closing Remarks</b>	

<b>5th Solar &amp; Stellar Astronomy Big Data (SABiD) -- Workshop on Management, Search, and Mining of Massive Repositories of Solar and Stellar Astronomy Data</b>		
<i>Workshop Chairs: Rafal A. Angryk, Piet C. Martens, Russel J. White, Dustin J. Kempton, Berkay Aydin</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
8:00-8:20	An Application of Spatio-temporal Co-occurrence Analyses for Integrating Solar Active Region Data from Multiple Reporting Modules	Xumin Cai
8:25-8:45	Window-Based Feature Extraction Method using XGBoost for Time Series Classification of Solar Flares	Renan Sauteraud
8:50-9:10	Solar Flare Prediction Using two-tier Ensemble with Deep Learning and Gradient Boosting Machine	Tommy Dang
9:15-9:35	A Deep Learning Model with Multi-Scale Skip Connections for Solar Flare Prediction Combined with Prior Information	Tian Han
9:35-9:55	<b>Coffee Break</b>	

10:00-10:20	Solar Flare Classification with Time Series Profiling	Ruizhe Ma
10:25-10:45	Solar event tracking with Deep Regression Networks: A proof of concept evaluation	Juan M Banda
10:50-11:10	Toward Filament Segmentation Using Deep Neural Networks	Azim Ahmadzadeh
11:15-11:35	Towards Understanding the Impact of Statistical Time Series Features for Flare Prediction Analysis	Dustin J. Kempton
<b>11:40-12:00</b>	<b>Closing Remarks</b>	

<b>Workshop on Big Data for Cybersecurity (BigCyber 2019)</b> <i>Workshop Chairs: Dr Karuna Joshi &amp; Dr. Bhavani Thuraisingham</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
8:00 to 8:10 am	<i>Opening Remarks</i>	Dr. Karuna Pande Joshi, Assistant Professor, Dept. of Information Systems, UMBC
8:10 to 8:30 am	<i>Extracting Rich Semantic Information about Cybersecurity Events</i>	Taneeya Satyapanich, Tim Finin, and Francis Ferraro
8:30 to 8:50 am	<i>An Extremely Lightweight Approach for DDoS Detection at Home Gateways</i>	Gabriel Mendonça, Gustavo Santos, Edmundo de Souza e Silva, Rosa Leão, Daniel Menasché, and Donald Towsley,
8:50 to 9:10 am	<i>Intelligent Feature Engineering for Cybersecurity</i>	Paul Maxwell, Elie Alhajjar, and Nathaniel Bastian
9:10 to 9:30 am	<i>The Effectiveness of Edge Centrality Measures for Anomaly Detection</i>	Candice Mitchell, Rajeev Agrawal, and Joshua Parker
9:30 to 9:50 am	<i>SecP2I : A Secure Multi-party Discovery of Personally Identifiable Information (PII) in Structured and Semi-structured Datasets</i>	Amine Mrabet, Mehdi Bentounsi, and Patrice Darmon
9:50 to 10:10 am	<i>Streaming Temporal Graphs: Subgraph Matching</i>	Eric Goodman and Dirk Grunwald
10:10 to 10:30 am	<i>Deepfake Detection and Challenges – A Study</i>	Md. Shohel Rana and Andrew H. Sung
10:30 to 10:45 am	<b>Coffee Break</b>	
10:45 am	<i>Keynote Speaker Introduction</i>	Dr. Rajeev Agrawal, PhD Computer Scientist, Information Technology Laboratory, Engineer Research and Development Center, U.S. Army Corps of Engineers
10:45 to 11:30 noon (Keynote)	<b><i>Exploring the role of Big Data in Defensive Cyber Deception and Adaptive Moving Target Defenses</i></b> <i>by</i> <b><i>Dr. Sunny James Fugate, Ph.D.,</i></b> <i>Senior Scientific Technical Manager for Cyber Warfare, Cyber / Science &amp; Technology Department, Naval Information Warfare Center, Pacific</i>	
11:30 to 11:50 pm	<i>Are We Really Protected? An Investigation into the Play Protect Service</i>	Shinelle Hutchinson, Bing Zhou, and Umit Karabiyik
11:50 to 12:10 pm	<i>Considering the Blackbox: An Investigation of Optimization Techniques with Completely Balanced Datasets of Packet Traffic</i>	Bruce Hartpence and Andres Kwasinski
<b>12:10 to 12:15 pm</b>	<b>Closing Remarks</b>	
<b>12:15 to 1:30 pm</b>	<b>Lunch</b>	

<b>The 2nd International Workshop on Big Media Dataset Construction, Management and Applications</b> <i>Workshop Chairs: Mingli Song, Mingyu You</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
13:30-13:50	Single Image Dehazing via Lightweight Multi-scale Networks	Guiying Tang, Li Zhao, Runhua Jiang, and Xiaoqin Zhang

13:50-14:10	Single-Image Dehazing Using Color Attenuation Prior Based on Haze-Lines	Qianru Wang, Li Zhao, Guiying Tang, Hanli Zhao, and Xiaoqin Zhang
14:10-14:30	Robust Basketball Player Tracking Based on a Hybrid Detection Grouping Framework for Overlapping Cameras	Kuan Hsien Wu, Wan Lun Tsai, Tse Yu Pan, and Min Chun Hu
14:30-14:50	Automatic Landmark Placement for Large 3D Facial Image Dataset	Jerry Wang, Shiao-fen Fang, Meie Fang, Jeremy Wilson, Noah Herrick, and Susan Walsh
14:50-15:10	Scalable Document Image Information Extraction with Application to Domain-specific Analysis	Yingbin Zheng, Shuchen Kong, Wanshan Zhu, and Hao Ye
15:10-15:30	Semantic Correlations Loss: Improving Model Interpretability for Multi-class Classification	Xuezhi Tong, Rui Wang, Xiaochun Cao, and Wenqi Ren
15:30-15:50	<b>Coffee Break</b>	
15:50-16:10	Weighted Focus-Attention Deep Network for Fine-grained Image Classification	Cong Zou, Rui Wang, Xiaochun Cao, and Feixiao Lv
16:10-16:30	Multi-View Subspace Clustering based on Tensor Schatten-p Norm	Yongli Liu, Xiaoqin Zhang, Guiying Tang, and Di Wang
16:30-16:50	Structural Dictionary Learning based on Non-convex Surrogate of $L_{\{2,1\}}$	Xiaoju Lu, Guiying Tang, Di Wang, Xiaoqin Zhang, and Jingjing Zheng
16:50-17:10	RnR: Extraction of Visual Attributes from Large-Scale Fashion Dataset	Sungjae Lee, Yeonji Lee, Junho Kim, and Kyungyong Lee
17:10-17:30	Banknotes Serial Number Coding Recognition	Ruru Xu, Xinli Min, Liandeng Su and Jungang An
17:30-17:50	Regression-based Face Pose Estimation with Deep Multi-modal Feature Loss	Yanqiu Wu, Chaoqun Hong, Jun Yu, and Liang Chen
17:50-18:00	<b>Closing Remarks</b>	

<b>Big Food and Nutrition Data Management and Analysis – BFNDMA 2019</b> <i>Tome Eftimov, Bibek Paudel, Barbara Koroušić Seljak</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
09:00 - 09:15	Welcome and Introduction	Tome Eftimov, Barbara Koroušić Seljak
09:15 - 09:50	FoodKG: A Semantics-Driven Knowledge Graph for Food Recommendation	Oshani Seneviratne
09:50 - 10:10	Comparing Semantic and Nutrient Value Similarities of Recipes	Gordana Ispirova
10:10 – 10:30	<b>Coffee Break</b>	
10:30 – 10:50	Optimization of arable land use towards meat-free and climate-smart agriculture: A case study in food self-sufficiency of Vietnam	Vladimir Kuzmanovski
10:50 – 11:10	Semi-Automatic Crowdsourcing Tool for Online Food Image Collection and Annotation	Zeman Shao
11:10 – 11:30	Exploring Dietary Intake Data collected by FPQ using Unsupervised Learning	Nina Reščič
11:30 – 12:05	Invited Talk 2	Karl Aberer
12:05 – 13:30	<b>Lunch</b>	
13:30 – 13:50	Food Waste Ontology: A Formal Description of Knowledge from the Domain of Food Waste	Riste Stojanov
13:50 – 14:10	From DIKW pyramid to graph database: a tool for machine processing of nutritional epidemiologic research data	Chen Yang
14:10 – 14:30	Exploring a standardized language for describing foods using embedding techniques	Gorjan Popovski
14:30 – 15:30	Light Talks (5 min presentation for each poster)	8 accepted papers
15:30 – 16:20	<b>Coffee Break</b>	
16:20 – 17:20	<b>Poster Presentation and Discussion</b>	All accepted papers
17:20 – 17:30	<b>Closing Remarks</b>	

<b>2019 International Workshop on IoT Big Data and Blockchain (IoTBB'2019)</b> <i>Workshop Chairs: Professor Huaglori Tianfield, and Professor Feng Qian</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
10:00-10:05 am	Opening Words	
10: 05-10:30 am	RF-MSiP: Radio Frequency Multi-source Indoor Positioning	Vishal Perekadan, Tathagata Mukherjee, Chaity Banerjee, and Eduardo Pasilliao

10:30-10:55 am	Resilient Activities Tracking in a Smart Home using Ultrasonic Sensors	Kashyap Venkatesh, Bashar Barmada, Veronica Liesaputra, and <u>Guillermo Ramirez-Prado</u>
10:55-11:20 am	IoT-based Multi-view Machine Vision Systems	Emmanuel Castillo, and <u>Ali Ahmadinia</u>
11:20-11:45 am	IBFRAME: IoT Data Processing Framework for Intelligent Building Management	<u>Dongwoo Kwon</u> , Kisu Ok, and Youngmin Ji
11:45-12:10 noon	Non-intrusive Behavior Awareness for Residents of a Smart House	<u>Guillermo Ramirez-Prado</u> , Bashar Barmada, and Veronica Liesaputra
04:20-04:45 pm	Detecting DoS Attack in Smart Home IoT Devices Using a Graph-Based Approach	Ramesh Paudel, Timothy Muncy, and William Eberle
04:45-05:10 pm	Centralized Trust Scheme for Cluster Routing of Wireless Sensor Networks	<u>Yunfan Li</u> , Nesrine Berjab, Hieu Hanh Le, and Haruo Yokota
05:10-05:35 pm	A Microservices Platform for Monitoring and Analysis of IoT Traffic Data in Smart Cities	Antonio De Iasio, Angelo Furno, Lorenzo Goglia, and <u>Eugenio Zimeo</u>
05:35-05:45 pm	Break	
05:45-06:10 pm	Blockchain and IoT for Delivery Assurance on Supply Chain (BIDAS)	Mehmet Demir, Ozgur Turetken, and Alexander Ferworn
06:10-06:35 pm	Privacy in IoT Blockchains: with Big Data comes Big Responsibility	Steven Wright
06:35-07:00 pm	Blockchain and Health Services	Huaglory Tianfield
07:00 pm	Closing Remarks	

The Third Workshop on Big Data for Economic and Business Forecasting <i>Workshop Chairs: Wei Shang, Matthew Harding, Xingfen Wang</i>		
Time	Title	Presenter/Author
9:00-9:20	Natural gas price prediction with big data	Wei Xu
9:20-9:40	Study on the relationship between house rent and people congestion by time in Tokyo based on mobile phone GPS data	Yinglan Qin
9:40-10:00	The Research on Cross-border Online Shopping Transaction Risk Based on Online Data Access	Xingfen Wang
10:00-10:20	Inventory Cost Control Model for Fresh Product Retailers Based on DQN	Ruoying SUN
Coffee Break		
10:40-11:00	Stock Index Forecasting by Hidden Markov Models with Trends Recognition	Wei Shang
11:20-11:40	A Conversational User Interface for Stock Analysis	Paula Lauren
11:40-12:00	What Make a Network Novel Popular? Implications for “Qidian.com”	Liangqiang Li
Closing Remarks		

Deep Graph Learning: Methodologies and Applications <i>Workshop Chairs: Lingfei Wu, Jiliang Tang, Liang Zhao, Tyler Derr</i>		
Time	Title	Presenter/Author
8:30 - 8:40am	Opening and Welcome	Lingfei Wu (Co-Chair), IBM Research, USA
8:40 - 9:25 am	Keynote 1	Berthold Reinwald, IBM Research, USA
9:25 - 10:10 am	Keynote 2	Yang Zhou, Auburn University, USA
10:10 - 10:30 am	Coffee Break	
10:30 - 11:15 am	Keynote 3	Dawei Zhou, University of Illinois Urbana-Champaign, USA
11:15 am - 12:00 pm	Keynote 4	Aditya Prakash, Virginia Tech, USA
12:00 - 12:30 pm	Poster Session	
Closing Remarks		

IEEE Workshop on Machine Learning for Big Data Analytics in Remote Sensing <i>Workshop Chair: Dr. Maryam Rahnemounfar</i>		
Time	Title	Presenter/Author

8:50-9:00	Welcome and Introduction	Maryam Rahnemoonfar
9:00-9:20	Remote Sensing Object Localization with Deep Heterogeneous Superpixel Features	Grant Scott
9:20-9:40	Novel Deep-Learning-Based Spatial-Spectral Feature Extraction for Hyperspectral Remote Sensing Applications	Bishwas Praveen
9:40-10:00	<b>Coffee Break</b>	
10:00-10:20	A two-stage framework for big spatial data processing to support disaster response	Jie Gong
10:20-10:40	Smart Tracking of Internal Layers of Ice in Radar Data via Multi-Scale Learning	Masoud Yari
10:40-11:00	Decision-Level Fusion of DNN Outputs for Improving Feature Detection Performance on Large-Scale Remote Sensing Image Datasets	Alan B Cannaday II
11:00-11:20	Scaling Deep Learning-Based Analysis of High-Resolution Satellite Imagery with Distributed Processing	Mai Nguyen
11:20-11:40	A Comparison of Deep Learning Vehicle Group Detection in Satellite Imagery	Grant Scott
11:40-12:00	Wavelet Features Greedy Clustering for Remotely Sensing Identification	Zhaohui Wang
12:00-12:10	<b>Closing Remarks</b>	

<b>The First Workshop on Security and Privacy on Blockchain</b> <i>Workshop Chairs: Raymond Choo, Pan Li, and Xiaodong Lin</i>		
Time	Title	Presenter/Author
10:30AM-12:00PM	<i>Bitcoin, the First Decentralized Cryptocurrency</i>	Mohamed Rasslan
12:00PM-12:30PM	<i>Asynchronous Blockchain-based Privacy-preserving Training Framework for Disease Diagnosis</i>	Xuhui Chen
12:30PM-1PM	<i>Privacy-Preserving Statistical Analysis of Health Data Using Paillier Homomorphic Encryption and Permissioned Blockchain</i>	Mahdi Ghadamyari
	<b>Closing Remarks</b>	

<b>The 3rd International Workshop on Big Data for Financial News and Data</b> <i>Workshop Chairs: Quanzhi Li, Xiaozhong Liu, Sameena Shah</i>		
Time	Title	Presenter/Author
8:40-9:00am	<i>Predicting the daily number of payment transactions in the largest bank in the Netherlands: Application to Banking Data</i>	Maartje Corstjens, Marzieh Bakhshandeh, Pinar Kahraman, and Joost Bosman
9:00-9:20am	<i>Deep Learning Approaches for Sentiment Analysis on Financial Microblog Dataset</i>	Savas Yildirim, Dhanya Jothimani, Can Kavaklioglu, and Ayse Basar
9:20-9:40am	<i>Deal or No Deal: Predicting Mergers and Acquisitions at Scale</i>	Ryan Moriarty, Howard Ly, Ellie Lan, and Suzanne McIntosh
9:40-10:00am	<i>A Framework of Applying Kelly Stationary Index to Stock Trading in Taiwan Market</i>	Jia-Hao Syu, Mu-En Wu, and Jan-Ming Ho
10:10-10:30am	<b>Coffee Break</b>	
10:30-10:50am	<i>Stock Prediction using Deep Learning and Sentiment Analysis</i>	Yichuan Xu and Vlado Keselj
10:50-11:10am	<i>FinDX: A Versatile, Low-Resource Approach to Financial Website Classification</i>	Alissa Ostapenko, Rodica Neamtu, and Frazer Anderson
11:10-11:30am	<i>Deep Learning for the Prediction of Stock Market Trends</i>	Arvand Fazeli and Sheridan Houghten
11:30-11:50am	<i>Structuring Time Series Data to Gain Insight into Agent Behaviour</i>	Najim Al-baghdadi, Wojciech Wisniewski, David Lindsay, Sian Lindsay, Yuri Kalnishkan, and Chris Watkins
12:00-1:30pm	<b>Lunch at your own</b>	
1:30-1:50pm	<i>Dimension Estimation of Equity Markets</i>	Nitish Bahadur, Randy Paffenroth, and Kelum Gajamannage
1:50-2:10pm	<i>Peer Firm Identification Using Word Embeddings</i>	Taeyoung Kee

2:10-2:30pm	<i>A Semi-Supervised Approach for Identification of the Sections in Charge of RFQ Documents</i>	yiou.wang@fujixerox.co.jp, izumo.hidetaka@fujixerox.co.jp
2:30-2:50pm	<i>KryptoOracle: A Real-Time Cryptocurrency Price Prediction Platform Using Twitter Sentiments</i>	Shubhankar Mohapatra, Nauman Ahmed, and Paulo Alencar
2:50-3:10pm	<i>Evaluating Sentiment Classifiers for Bitcoin Tweets in Price Prediction Task</i>	Ahmed Balfagih and Vlado Keselj,
3:10-3:30pm	<i>CoStock: A DeepFM Model for Stock Market Prediction with Attentional Embeddings</i>	Jieyun Huang, Xi Zhang, and Binxing Fang
3:30pm	<b>Closing Remarks</b>	
<b>3:40-4:00pm</b>	<b>Coffee Break</b>	

<b>4th International Workshop on Big Data Transfer Learning (BDTL)</b> <i>Workshop Chairs: Ming Shao, and Yun Fu</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
1:30 PM--2:00 PM	Opening Remarks	Workshop co-chair
2:00 PM--2:25 PM	Decoder Transfer Learning for Predicting Personal Exposure to Air Pollution	PEIJIANG ZHAO and Koji Zettsu
2:25 PM--2:50 PM	Advertiser-Assisted Behavioral Ad-Targeting via Denoised Distribution Induction	Kei Yonekawa, Niu Hao, Mori Kurokawa, Arei Kobayashi, Daichi Amagata, Takuya Maekawa, and Takahiro Hara
2:50 PM--3:40 PM	Keynote Speech	TBD
3:40--4:10	<b>Coffee Break</b>	
4:10 PM--4:35 PM	DC2: A Divide-and-conquer Algorithm for Large-scale Kernel Learning with Application to Clustering	Ke Alexander Wang, Xinran Bian, Pan Liu, and Donghui Yan
4:35 PM--5:00 PM	On Online Hate Speech Detection. Effects of Negated Data Construction	Mourad Oussalah and Abderrauof Cheniki
5:00 PM--5:25 PM	Multi-View, Generative, Transfer Learning for Distributed Time Series Classification	Sreyasee Das Bhattacharjee, William J. Tolone, Ashish Mahabal, Mohammed Elshambakey, Isaac Cho, Abdullah al-Raihan Nayeem, Junsong Yuan, and George Djorgovski
<b>5:25 PM--5:30 PM</b>	<b>Closing Remarks</b>	

<b>The next frontier of big data from LiDAR</b> <i>Workshop Chairs: Wang Zhou, IBM Research</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
1:30 – 1:55	<i>Learning and Recognizing Archeological Features from LiDAR Data</i>	Marcus Freitag
1:55 – 2:20	<i>Efficient LiDAR Point Cloud Data Encoding for Scalable Data Management within the Hadoop Eco-system</i>	Debra F. Laefer
2:20 – 2:45	<i>Defense-PointNet: Protecting PointNet Against Adversarial Attacks</i>	Yu Zhang
2:45 – 3:10	<i>N-dimensional Geospatial Data and Analytics for Critical Infrastructure Risk Assessment</i>	Levente Klein
<b>3:10</b>	<b>Closing Remarks</b>	

<b>Streaming Systems and Realtime Machine Learning (STREAM-ML) Workshop</b> <b>Tuesday, December 10, 2019</b> <b>Echo Park</b>			
<b>Time</b>	<b>Title</b>	<b>Presenter*/Authors list</b>	<b>Affiliations</b>
<b>4:20 pm</b>	<b>Welcome</b> <i>Judy Qiu, Geoffrey Fox and Madhav Marathe</i>		

<b>4:20 pm</b>	Performance Characterization and Modeling of Serverless and HPC Streaming Applications	Andre Luckow* Shantenu Jha	BMW Group, Germany Rutgers University, USA
<b>4:40 pm</b>	Streaming Machine Learning Algorithms with Big Data Systems	Vibhatha Abeykoon*, Supun Kamburugamuve, Kannan Govindarajan, Pulasthi Wickramasinghe, Chathura Widanage, Niranda Perera, Ahmet Uyar, Gurhan Gunduz, Selahattin Akkas, Gregor Von Laszewski	Indiana University, USA
<b>5:00 pm</b>	Benchmarking Deep Learning for Time Series: Challenges and Directions	Xinyuan Huang*, Geoffrey Fox, Sergey Serebryakov, Ankur Mohan, Pawel Morkisz, Debojyoti Dutta	Cisco Systems, USA Indiana University, USA Hewlett Packard Enterprise, USA In-Q-Tel, USA AGH University of Science and Technology, Poland
<b>5:20 pm</b>	A Fast Video Image Detection using TensorFlow Mobile Networks for Racing Cars	Selahattin Akkas*, Sahaj Singh Maini, Judy Qiu	Indiana University, USA
<b>5:40 pm</b>	MATRICES: A System for Human-Machine Hybrid Forecasting of Geopolitical Events	David Huber, Nigel Stepp, Aruna Jammalamadaka, Tiffany Kim, Sam Johnson, Dana Warmesley, Tsai-Ching Lu*	HRL Laboratories, USA
<b>6:00 pm</b>	DeepLite: Real-Time Deep Learning Framework for Neighborhood Analysis	Duy Ho, Raj Marri*, Sirisha Rella Yugyung Lee	University of Missouri - Kansas City, USA
<b>6:20 pm</b>	Adaptive Hoeffding Tree with Transfer Learning for Streaming Synchrophasor Data Sets	Zakaria El Mrabet*, Daisy Flora Selvaraj, Prakash Ranganathan	University of North Dakota, USA
<b>6:40 pm</b>	<b>Discussions and Closing Remarks</b>		



Room: Mt Washington		
Time	Title	Presenter/Author
08:00am – 08:05am	<b>Session PSBD19_1: Opening</b> <b>Chair: Alfredo Cuzzocrea</b>	
08:05am – 08:45am	<b>Session PSBD19_2: Malware and Intrusion Detection</b> <b>Chair: Alfredo Cuzzocrea</b>	
08:05am – 08:25am	Droid-NNet: Deep Learning Neural Network for Android Malware Detection	Mohammad Masum and Hossain Shahriar
08:25am – 08:45am	The Effectiveness of Learning Trees on Network Intrusion Detection	Dan Lo
10:15am – 12:10am	<b>Session PSBD19_3: Cybersecurity and Privacy-Preserving Data Mining</b> <b>Chair: Alfredo Cuzzocrea</b>	
10:15am – 10:35am	The Anti-Data-Mining (ADM) Framework - Better Privacy on Online Social Networks and Beyond	Shah Mahmood
10:35am – 10:55am	Privacy and Security of Big Data in AI Systems: A Research and Standards Perspective	Saharnaz Dilmaghani, Matthias R. Brust, Gregoire Danoy, Natalia Cassagnes, Johnatan Pecero, and Pascal Bouvry
10:55am – 11:15am	Utility and Privacy Assessments of Synthetic Data for Regression Tasks	Markus Hittmeir, Andreas Ekelhart, and Rudolf Mayer
11:15am – 11:35am	Adversarial Training for Privacy-Preserving Deep Learning Model Distribution	Mohammed Alawad
11:35am – 11:55am	Experiential Learning: Case Study-based Portable Hands-on Regression Labware for Cyber Fraud Prediction	Hossain Shahriar, Michael Whitman, Dan Lo, Fan Wu, Cassandra Thomas, and Alfredo Cuzzocrea
11:55am – 12:15am	Deep Neural Networks as Similitude Models for Sharing Big Data	Philip Derbeko, Shlomi Dolev, and Ehud Gudes
12:15am - 2:00pm	<b>Lunch</b>	
04:20pm – 06:00pm	<b>Session PSBD19_4: Privacy-Preserving Big Data Management</b> <b>Chair: TBA</b>	
04:20pm – 04:40pm	An Identity Privacy Preserving IoT Data Protection Scheme for Cloud Based Analytics	Christian Gehrman and Martin Gunnarsson
04:40pm – 05:00pm	Privacy-preserving Top-k Dominating Queries in Distributed Multi-party Databases	Mahboob Qaosar, Kazi Md. Rokibul Alam, Chen Li, and Yasuhiko Morimoto
05:00pm – 05:20pm	Local Differential Privacy with K-anonymous for Frequency Estimation	Dan Zhao, Hong Chen, Suyun Zhao, Cuiping Li, Xiaoying Zhang, and Ruixuan Liu
05:20pm – 05:40pm	Kratos: A secure, authenticated and publicly verifiable system for educational data using the blockchain	Velislava Hillman and Varunram Ganesh
05:40pm – 06:00pm	Privacy-Preserving Multi Keyword Search on Encrypted Outsourced Data	Carson Leung, Bryan Wodi, and Alfredo Cuzzocrea
06:00pm – 07:20pm	<b>Session PSBD19_5: Privacy-Preserving Big Data Processing and Analysis</b> <b>Chair: TBA</b>	
06:00pm – 06:20pm	Improving k-Nearest Neighbor Pattern Recognition Models for Privacy-Preserving Data Analysis	Walisa Romsaiyud, Henning Schnoor, and Wilhelm Hasselbring
06:20pm – 06:40pm	Distributed Consensus Reduced Support Vector Machine	Hsiang-Hsuan Chen and Yuh-Jye Lee
06:40pm – 07:00pm	Anonymous Privacy-Preserving Scheme for Big Data Over the Cloud	Zeyad Al-Odat and Samee Khan
07:00pm – 07:20pm	RIBS: Risky Blind-Spots for Attack Classification Models	Mikel Joaristi, Arthur Putnam, Alfredo Cuzzocrea, and Edoardo Serra



## Special Symposia

TBD

# Special Sessions

## 5<sup>th</sup> SPECIAL SESSION ON INTELLIGENT DATA MINING

**Session Organizer:** Uraz YAVANOGLU, PhD

### Summary:

After the successes of the first, second, third and fourth editions of Special Session on Intelligent Data Mining in Santa Clara, CA (2015), Washington, DC (2016), Boston, MA (2017), Seattle, WA, (2018) and the fifth Special Session on Intelligent Data Mining in Los Angeles, CA (2019) will continue promoting and disseminating the knowledge concerning several topics and technologies related to data mining science.

Artificial Intelligence (AI) & Machine Learning (ML) fields are interdisciplinary, including computer science, mathematics, psychology, linguistics, philosophy, neuroscience etc. This interdisciplinary special session seeks scientific understanding on data and intelligence.

This session may help to create scientific evolution to propose robust and powerful schemes between human nature and big data processing.

Intelligent Data Mining session open to every researcher as well as industrial partners,

### Short Bio.

Uraz Yavanoglu was born in Ankara, capital city of Turkey. He is an assistant professor at Gazi University Department of Computer Engineering and holding vice-chairman position. 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 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-10 December 2019 Tuesday		
Time	Title	Presenter/Author
07:00am-08:00am		<b>Registration</b>
08:00am-08:10am		<b>Session Keynote Speech</b> Uraz Yavanoglu, PhD
08:10am-08:20am	SP01233	Lukasz Korycki, Alberto Cano, and Bartosz Krawczyk, <i>Active Learning with Abstaining Classifiers for Imbalanced Drifting Data Streams</i>
08:20am-08:30am	SP01231	William Sleeman and Bartosz Krawczyk <i>Bagging Using Instance-Level Difficulty for Multi-Class Imbalanced Big Data Classification on Spark</i>
08:30am-08:40am	SP01201	Izzat Alsmadi, Muhammad Al-abdullah, and Hisham Alsmadi <i>Popular Search Terms and Stock Price Prediction</i>
08:40am-08:50am	SP01239	Orhun Bugra Baran, Saim Sunel, Pinar Karagoz, and Ismail Hakki Toroslu <i>ATM Withdrawal Amount Forecasting Through Neural Architectures</i>
08:50am-09:00am	SP01206	Vinnie Ko, Stefan Oehmcke, and Fabian Gieseke <i>Magnitude and Uncertainty Pruning Criterion for Neural Networks</i>
09:00am-09:10am	SP01221	Stefan Oehmcke, Christoffer Thrysøe, Andreas Borgstad, Marcos Antonio Vaz Salles, Martin Brandt, and Fabian Gieseke <i>Detecting Hardly Visible Roads in Low-Resolution Satellite Time Series Data</i>

09:10am-09:20am	SP01207	Nora Alkhamees and Maria Fasli <i>The Dynamic-FPM: An Approach for Identifying Events From Social Networks Using Frequent Pattern Mining and Dynamic Support Values</i>
09:20am-09:30am	SP01210	Ragini Kihlman and Maria Fasli <i>Augmenting co-training with recommendations to classify human right violations</i>
09:30am-09:40am	SP01204	Bob Vanderheyden, Ying Xie, and Mohan Rachumallu <i>Net Promoter Sentiment Classifier Using OHPL-ALL</i>
09:40am-10:00am	<b>Break</b>	
10:00am-10:10am	SP01213	Paolo Rosso, Dingqi Yang, and Philippe Cudré-Mauroux <i>Revisiting Text and Knowledge Graph Joint Embeddings: The Amount of Shared Information Matters!</i>
10:10am-10:20am	SP01214	Merve Astekin, Selim Özcan, and Hasan Sözer <i>Incremental Analysis of Large-Scale System Logs for Anomaly Detection</i>
10:20am-10:30am	SP01215	Buvaneswari Ramanan, Lawrence Drabeck, Thomas Woo, Troy Cauble, and Anil Rana <i>Eliminating Data Collection Bottleneck for Wake Word Engine Training Using Found and Synthetic Data</i>
10:30am-10:40am	SP01216	Hannu Reittu, Ville Kotovirta, Lasse Leskelä, Hannu Rummukainen, and Tomi Rätty <i>Towards analyzing large graphs with quantum annealing</i>
10:40am-10:50am	SP01218	Minh-Son Dao, Ngoc-Thanh Nguyen, and Koji Zettsu <i>Multi-time-horizon Traffic Risk Prediction using Spatio-Temporal Urban Sensing Data Fusion</i>
10:50am-11:00am	SP01217	Manal Almuammar and Maria Fasli, <i>Deep Learning for Non-stationary Multivariate Time Series Forecasting</i>
11:00am-11:10am	SP01220	Saratchandra Indrakanti, Svetlana Strunjas, Shubhangi Tandon, and Manojkumar Kannadasan <i>Influence of Neighborhood on the Preference of an Item in eCommerce Search</i>
11:10am-11:20am	SP01225	Gulustan Dogan <i>Sentiment Analysis in Turkish with Deep learning</i>
11:20am-11:30am	SP01226	Gulustan Dogan <i>A Data-Driven Approach to Kinematic Analytics of Spinal Motion</i>

11:30am-11:40am	SP01229	Sanchita Basak, Abhishek Dubey, and Leao Bruno <i>Analyzing the Cascading Effect of Traffic Congestion Using LSTM Networks</i>
11:40am-11:50am	SP01230	Andres Leiva-Araos, Denis Khryashchev, Hector Allende-Cid, and Huy T. Vo <i>Tackling the Neighboring Network Hit Problem in Cellular Data</i>
11:50am-12:00pm	SP01232	Abdelouahab Khelifati, Mourad Khayati, and Philippe Cudré-Mauroux <i>CORAD: Correlation-Aware Compression of Massive Time Series using Sparse Dictionary Coding</i>
12:00pm-12:10pm	SP01235	Junlan Lu, Nikhil Takappa Saunshi, Aldrich Mangune, Magdalini Eirinaki, Bin Yu, and Cricket Liu <i>A SERP Mining Approach for Topic Classification of DNS Requests</i>
12:10pm-1:00pm	<b>Lunch Break</b>	
1:00pm-1:10pm	SP01236	Chandadevi Giri, Ulf Johansson, and Tuwe Löfström <i>Predictive Modeling of Campaigns to Quantify Performance in Fashion Retail Industry</i>
1:10pm-1:20pm	SP01208	Qinghan Xue, Abhishek Kolagunda, Steven Eliuk, and Xiaolong Wang <i>AWDF: An Adaptive Weighted Deep Fusion Architecture for Multi-modality Learning</i>
1:20pm-1:30pm	BigD319	Wei Quan, Jinli Zhang, and Xiaohua Hu <i>End-to-End Joint Opinion Role Labeling with BERT</i>
1:30pm-1:40pm	BigD324	Shuaidong Pan, Tianran Hu, Shujing Sun, Jianbo Yuan, and Jiebo Luo <i>Help Oneself in Helping the Others: the Ecology of Online Support Groups</i>
1:40pm-1:50pm	SP01222	Jabir Alshehabi Al-Ani and Maria Fasli, <i>Probabilistic Named Entity Recognition for non-standard format entities using co-occurrence word embeddings</i>
1:50pm-2:00pm	SP01223	Flavio Giobergia and Elena Baralis, <i>Fast Self-Organizing Maps Training</i>
02:00pm-02:10pm	BigD333	Jiang Bian, Weibo Wang, Xiang Zhang, Wei Wang, Arthur Huang, and Zhishan Guo <i>On Generating Dominators of Customer Preferences</i>
02:10pm-02:20pm	BigD365	Yujing Chen and Huzefa Rangwala <i>Attention-based Multi-task Learning for Sensor Analytics</i>
02:20pm-02:30pm	BigD411	Yasmeen George, Shanika Karunasekera, Aaron Harwood, and Kwan Hui Lim <i>Spatio-temporal Event Detection using Poisson Model and Quad-tree on Geotagged Social Media</i>

02:30pm-02:40pm	BigD414	Huaming Chen, Lei Wang, Yaochu Jin, Chi-Hung Chi, Fucun Li, Huaiyuan Chu, and Jun Shen <i>Hyperparameter Estimation in SVM with GPU Acceleration for Prediction of Protein-Protein Interactions</i>
02:40pm-02:50pm	BigD430	Idir Benouaret, Sihem Amer-Yahia, Christiane Kamdem Kengne, and Jalil Chagraoui <i>A Bi-Objective Approach for Product Recommendations</i>
02:50pm-03:00pm	BigD515	Kurosh Madani, Antonio M. Rinaldi, and Cristiano Russo <i>Merging Large Ontologies using BigData GraphDB</i>
03:00pm-03:10pm	BigD520	Artem Lutov, Mourad Khayati, and Philippe Cudré-Mauroux <i>DAOC: Stable Clustering of Large Networks</i>
03:10pm-03:20pm	BigD582	Mariappan Asokan <i>A robust, efficient, and balanced parallel algorithm for finding connected components</i>
03:20pm-03:30pm	BigD607	Ramoza Ahsan, Muzammil Bashir, Rodica Neamtu, Elke Rundensteiner, and Gabor Sarkozy <i>Nearest Neighbor Subsequence Search in Time Series Data</i>
03:30pm-03:40pm	BigD647	Cansu Sen, Thomas Hartvigsen, Xiangnan Kong, and Elke Rundensteiner <i>Learning Temporal Relevance in Longitudinal Medical Notes</i>
03:40pm-03:50pm	BigD654	Yi Feng, Yi Zhou, and Vahid Tarokh <i>Recurrent Neural Network-Assisted Adaptive Sampling for Approximate Computing</i>
03:50pm-04:00pm	BigD655	Yang Zhou, Yan Huang, Joseph McGlynn, and Alexander Han <i>Trust Inference for Rideshare through Co-training on Social Media Data</i>
04:00pm-04:20pm	<b>Break</b>	
04:20pm-04:30pm	SP01224	Jeeyung Kim <i>Time-varying Item Feature Conditional Variational Autoencoder for Collaborative Filtering</i>
04:30pm-04:40pm	BigD659	Aaron Harwood, Shanika Karunasekera, Michelle Vanni, Lucia Falzon, Prarthana Padia, and Amila Silva <i>Understanding Multilingual Communities through Analysis of Code-switching Behaviors in Social Media Discussions</i>
04:40pm-04:50pm	BigD696	Zerong Liu and Aidong Lu <i>Explainable Visualization for Interactive Exploration of CNN on Wikipedia Vandal Detection</i>
04:50pm-05:00pm	BigD699	Yunzhe Fang, Xiao-Yang Liu, and Hongyang Yang <i>Practical Machine Learning Approach to Capture the Scholar Data Driven Alpha in AI Industry</i>

05:00pm-05:10pm	SP01211	Pooja Bhagat, Aparna Varde, and Anna Feldman <i>WordPrep: Word-based Preposition Prediction Tool</i>
05:10pm-05:20pm	SP01227	Reham Alamro and Abdou Youssef <i>Effects of Data Reduction Methods and Rates on Classifiers</i>
05:20pm-05:30pm	SP01234	Thankgod Obasi and M. Omair Shafiq, <i>Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attack and Diseases</i>
05:30pm-05:40pm	BigD693	Wenhao Zhang, Ramin Ramezani, and Arash Naeim <i>WOTBoost: Weighted Oversampling Technique in Boosting for imbalanced learning</i>
05:40pm-05:50pm	SP01219	Piyush Yadav and Edward Curry <i>VidCEP: Complex Event Processing Framework to Detect Spatiotemporal Patterns in Video Streams</i>
05:50pm-06:00pm	SP01237	Beyza Bagiroz, Metehan Guzel, Uraz Yavanoglu, and Suat Özdemir <i>QoS Prediction Methods in IoT : A Survey</i>
06:00pm-06:10pm	SP01238	İbrahim Kök, Burak Han Çorak, Uraz Yavanoglu, and Suat Özdemir <i>Deep Learning based Delay and Bandwidth Efficient Data Transmission in IoT</i>
06:10pm-06:20pm	SP01241	Kivanc Bayraktar, Uraz Yavanoglu, and Alper Ozbilen <i>A Rule-Based Holistic Approach for Turkish Aspect-Based Sentiment Analysis</i>
06:20pm-06:30pm	SP01242	Burak Ozcakmak, Alper Ozbilen, Uraz Yavanoglu, and Kübra Cin <i>Neural and Quantum Cryptography in Big Data: A Review</i>
06:30pm	Session Closing	Session Closing

### Special Session on Information Granulation in Data Science and Scalable Computing

Special Session on Information Granulation in Data Science and Scalable Computing		
Time	Title	Presenter/Author
10:05-10:30	AN EFFECTIVE AND SCALABLE DATA MODELING FOR ENTERPRISE BIG DATA PLATFORM (SP03217)	JAYESH PATEL
10:30-10:55	Study of the Effects of Visual Complexity and Consumer Experience on Visual Attention and Purchase Behavior through the Use of Eye Tracking (SP03213)	Ken Ishibashi, Chen Xiao, and Katsutoshi Yada
10:55-11:20	Mining Temporal Fuzzy Utility Itemsets by Tree Structure (SP03203)	Tzung-Pei Hong, Cheng-Yu Lin, Wei-Ming Huang, Shu-Min Li, Shyue-Liang Wang, and Jerry Chun-Wei Lin
11:20-11:45	Mining High-Utility Sequential Patterns from Big Datasets (SP03202)	Jerry Chun-Wei Lin, Yuanfa Li, Philippe Fournier-Viger, Youcef Djenouri, and Shyue-Liang Wang

11:45-12:10	Mining frequent temporal patterns from medical data based on fuzzy ranged relations (SP03212)	Shoji Hirano and Shusaku Tsumoto
	<b>Coffee Break</b>	
14:30-14:55	Bridging the Gap between Community and Node Representations: Graph Embedding via Community Detection (SP03206)	Artem Lutov, Dingqi Yang, and Philippe Cudr ̄̄-Mauroux
14:55-15:20	Coarse Graining of Data via Inhomogeneous Diffusion Condensation (SP03210)	Nathan Brugnone, Alex Gonopolski, Mark Moyle, Manik Kuchroo, David van Dijk, Kevin Moon, Daniel Colon-Ramos, Guy Wolf, Matthew Hirn, and Smita Krishnaswamy
15:20-15:45	Finding Archetypal Spaces Using Neural Networks (SP03209)	David van Dijk, Daniel Burkhardt, Matthew Amodio, Alexander Tong, Guy Wolf, and Smita Krishnaswamy
15:45-16:10	Radically Simplifying Gated Recurrent Architectures Without Loss of Performance (SP03211)	Jonathan Boardman and Ying Xie
	<b>Coffee Break</b>	
16:30-16:55	A GA-based Framework for Mining High Fuzzy Utility Itemsets (SP03201)	Jimmy Ming-Tai Wu, Jerry Chun-Wei Lin, Philippe Fournier-Viger, Tomasz Wiktorski, Tzung-Pei Hong, and Matin Pirouz
16:55-17:20	Utility-Driven Mining of High Utility Episodes (SP03205)	Wensheng Gan, Jerry Chun-Wei Lin, Han-Chieh Chao, and Philip S. Yu
17:20-17:45	Approximate Decision Tree Induction over Approximately Engineered Data Features (SP03215)	Agnieszka Chadzyska-Krasowska and Dominik Slezak
17:45-18:10	Estimation of Disease Code from Electro Patient Records (SP03214)	Shusaku Tsumoto, Tomohiro Kimura, and Shoji Hirano
	<b>Closing Remarks</b>	

<b>Special Track on FML</b>		
Tuesday, December 10th, 2019 - Location: Santa Barbara C		
16:20 - 16:25	Open Address	Prof. Qiang Yang
15 mins	Federated Learning with Bayesian Differential Privacy	Aleksei Triastcyn and Boi Faltings
15 mins	SGNN: A Graph Neural Network Based Federated Learning Approach by Hiding Structure	Guangxu Mei, Ziyu Guo, Shijun Liu, and Li Pan
15 mins	Measure Contribution of Participants in Federated Learning	Guan Wang, Charlie Xiaoqian Dang, and Ziyue Zhou
15 mins	Profit Allocation for Federated Learning	Tianshu Song, Yongxin Tong, and Shuyue Wei
5 mins	Break	

15:30-15:45	Secure and Efficient Federated Transfer Learning	Shreya Sharma, Xing Chaoping, Yang Liu, and Yan Kang
15 mins	Infer Latent Privacy for Attribute Network in Knowledge Graph	Zeyuan Cui, Li Pan, Shijun Liu, and Lizhen Cu
15 mins	Privacy-preserving Heterogeneous Federated Transfer Learning	Dashan Gao, Yang Liu, Anbu Huang, Ce Ju, Han Yu, and Qiang Yang
15 mins	Power Demand Response Incentive Pricing Model	Kun Zhang, Yuliang Shi, Yuecan Liu, and Zhongmin Yan,
18:30	Closing	

IEEE Big Data 2019 – 2 <sup>nd</sup> Special Session on HealthCare Data		
<i>Special Session Chairs: A. Teoman Naskali</i>		
Time	Title	Presenter/Author
08:00-08:10	Welcome	
08:10-08:30	The SERUMS tool-chain: Ensuring Security and Privacy of Medical Data in Smart Patient-Centric Healthcare Systems	Vladimir Janjic
08:30-08:50	A hybrid model using LSTM and decision tree for mortality prediction and its application in provider performance evaluation	Peichang Shi
08:50-09:10	Classification Models and Survival Analysis for Prostate Cancer Using RNA Sequencing and Clinical Data	Md Faisal Kabir
09:10-09:30	Using hospital administrative data to infer patient-patient contact via the consistent co-presence algorithm	Jeffrey Lienert
09:30-09:50	Enhancing Clinical Information Retrieval through Context-Aware Queries and Indices	Jungwei Fan
09:50-10:10	Regional Analysis of Death Rate due to Air Pollution in Turkey and its Neighbors	A. Teoman Naskali
10:10-10:30	<b>Coffee Break</b>	
10:30-10:50	Single-cell regulatory network inference and clustering from high-dimensional sequencing data	Aristidis Vrahatis
10:50-11:10	Analyzing Public Outlook towards Vaccination using Twitter	Rutuja Mahajan
11:10-11:30	The Development of Machine Learning Infused Outpatient Prognostic Models for tackling Impacts of Climate Change and ensuring Delivery of Effective Population Health Services	Chandrasekar Vuppalapati
11:30-11:50	A Data-Driven Approach for Continuous Adherence Predictions in Sleep Apnea Therapy Management	Matheus Araujo
11:50-12:00	<b>Closing Remarks</b>	



# BigData Cup Challenges

5th Solar & Stellar Astronomy Big Data (SABiD) -- Workshop on Management, Search, and Mining of Massive Repositories of Solar and Stellar Astronomy Data		
Workshop Chairs: Rafal A. Angryk, Piet C. Martens, Russel J. White, Dustin J. Kempton, Berkay Aydin		
Time	Title	Presenter/Author
8:00-8:20	An Application of Spatio-temporal Co-occurrence Analyses for Integrating Solar Active Region Data from Multiple Reporting Modules	Xumin Cai
8:25-8:45	Window-Based Feature Extraction Method using XGBoost for Time Series Classification of Solar Flares	Renan Sauteraud
8:50-9:10	Solar Flare Prediction Using two-tier Ensemble with Deep Learning and Gradient Boosting Machine	Tommy Dang
9:15-9:35	A Deep Learning Model with Multi-Scale Skip Connections for Solar Flare Prediction Combined with Prior Information	Tian Han
9:35-9:55	<b>Coffee Break</b>	
10:00-10:20	Solar Flare Classification with Time Series Profiling	Ruizhe Ma
10:25-10:45	Solar event tracking with Deep Regression Networks: A proof of concept evaluation	Juan M Banda
10:50-11:10	Toward Filament Segmentation Using Deep Neural Networks	Azim Ahmadzadeh
11:15-11:35	Towards Understanding the Impact of Statistical Time Series Features for Flare Prediction Analysis	Dustin J. Kempton
11:40-12:00	<b>Closing Remarks</b>	

BigData Cup Challenges - Suspicious Network Event Recognition		
Workshop Chairs: Dominik Ślęzak and Andrzej Janusz		
Time	Title	Presenter/Author
8:45 – 9:10	<i>AI @ Security On-Demand – Now and Future</i>	Joel Holland
9:10 – 9:35	<i>IEEE BigData 2019 Cup: Suspicious Network Event Recognition</i>	Andrzej Janusz
9:35 – 10:00	<i>Gradient boosting decision trees for cyber security threats detection based on network events logs</i>	Amy Ling Cen
10:00 – 10:25	<i>An Approach For Scale Suspicious Network Events Detection</i>	Cong Dong
10:25 – 10:45	<b>Coffee Break</b>	
10:45 – 11:10	<i>Identifying Truly Suspicious Events and False Alarms Based on Alert Graph</i>	Chen Zhang
11:10 – 11:30	<i>Application of XGBoost to the cyber-security problem of detecting suspicious network traffic events</i>	Łukasz Podlowski
11:30 – 11:50	<i>Automated Event Prioritization for Security Operation Center using Deep Learning</i>	Issa Traore
11:50 – 12:10	<i>IEEE BigData 2019 Cup: Binary Classification via Tensor Completion</i>	Teresa Ranadive
12:15 – 13:30	<b>Lunch Break</b>	
13:30 – 13:50	<i>An Apriori-based Data Analysis on Suspicious Network Event Recognition</i>	Hiroshi Sakai
13:50 – 14:10	<i>Suspicious Network Event Recognition Using A Modified Stacking Ensemble Machine Learning</i>	Angus F.M. Huang
14:10 – 14:30	<i>An Ensemble Approach for Suspicious Traffic Detection from High Recall Network Alerts</i>	Peilin Wu
14:30 – 14:50	<i>Suspicious Network Event Recognition Leveraging on Machine Learning</i>	Daniele Sartiano
14:50 – 15:10	<i>Naive Transfer Learning Approaches for Suspicious Event Prediction</i>	Chang Lin
15:10 – 15:30	<i>Models and Features with Covariate Shift Adaptation for Suspicious Network Event Recognition</i>	Shu-Yi Xie
15:30 – 15:35	<b>Closing Remarks</b>	
15:35 – 15:55	<b>Coffee Break</b>	

Multilingual Communities Workshop		
Chairs: Aaron Harwood, Shanika Karunasekera, Michelle Vanni		

Date/Time	Title	Presenter/Author
9 Dec 19/1330	Problem and Competition Overview	Shanika Karunasekera, University of Melbourne Michelle Vanni, Army Research Laboratory
1345	Anuj@IEEE BigData2019: A Novel Code-Switching Behavior Analysis in Social Media Discussions Natural Language Processing	Anuj Saini/Anuj Saini, Publicis Sapient
1415	Language Identification and Context-based Analysis of Code-switching Behaviors in Social Media Discussions	Akankshya Mishra/Akankshya Mishra & Yashvardhan Sharma
1445	Understanding Multilingual Communities through Analysis of Code-Switching Behaviors in Social Media Discussions	Shanika Karunasekera/Aaron Harwood, S. Karunasekera, M. Vanni, L. Falzon, P. Padia, Amila Silva
1515	Closing Remarks and Discussion	



Brain Data Bank Challenge - IEEE Brain Initiative (send questions to: [BDB-LA@IEEE.org](mailto:BDB-LA@IEEE.org)) - at the Big Data Conference 2019 – The Westin Bonaventure Hotel & Suites, 404 South Figueroa Street, Los Angeles, California, USA

Chair: Dr. Nan Chu



Keynote: Bruce Hecht



Keynote: Brent Lunceford



#### **Agenda - December 10, 2019**

08:30 - Big Data Conference Opening and Keynote, Rooms: San Francisco/San Jose, Sacramento

09:45 - *Coffee Break, Room: California Foyer*

#### **BDB Challenge Room: San Fernando**

10:00 - Brain Data Bank Challenge - Introduction and Ordering of Team Presentations

11:00 – Challenge Opening Keynote: “Emerging Capabilities in Accessing Brain Activity & the Future of Brain Data” – Bruce A. Hecht, Analog Devices, and IEEE Sensors Council, Boston, MA, USA.

12:00 – *Conference Lunch, Rooms: San Francisco/San Jose, Sacramento*

13:00 - Challenge Team Presentations

15:00 – Keynote: Neuroscience-based Entrepreneurship – Brent Lunceford, Memstronics, Austin, TX, USA.

- “Big Data and Neurotechnology – What could be ready for entrepreneur's undertaking?”
- Remarks for each team's presentation from a start-up perspective.

16:00 – *Coffee Break, Room: California Foyer*

16:20 – Challenge Closing Remarks: “Brain Data vs. Artificial Intelligence”

**December 11, 2019, Rooms: San Francisco/San Jose, Sacramento**

19:00 – **Winners' Awards** Announced at the Conference Banquet

## Panel

### **Addressing Big Data Heterogeneity Challenges: Recent Advances and Challenges Panel Session- IEEE Big Data Conf. 2019**

The real-world big data are largely unstructured, interconnected, and dynamic, in a variety of forms, including natural language text. Modern computers have demonstrated their tremendous power on search and reasoning on structured data. However, DBMSs require all data to be under the control of a single administrative domain and to conform to a single schema. In return for these limitations, a DBMS is able to provide rich data manipulation and query processing with well-understood, strong semantics.

Even though there are some recent big data software, such as sparkQL, and emerging querying mechanisms, like NewSQL, these are mainly designed to address enterprise data, which consists mostly of unstructured text and structured databases.

In data management scenarios today, it is rarely the case that all the data can be fit nicely into a conventional relational DBMS, or into any other single data model or system. Instead, developers are more often faced with a set of loosely connected data sources and thus must individually and repeatedly address low-level data management challenges across heterogeneous collections. These scenarios arise in enterprises (large or small): within and across government agencies, large science-related collaborations, libraries (digital or otherwise), battlefields, in “smart” homes, and even on one’s PC desktop or other personal devices.

The challenges include providing search and query capability; enforcing rules, integrity constraints, naming conventions, etc.; tracking lineage; providing availability, recovery, and access control; and managing evolution of data and metadata. Another key challenge to enable machine intelligence is to transform massive unstructured big data into structured knowledge. Furthermore, Big data over rich medias such as video, audio, photos are reality today. The consumption of these big data on commercial hardware is becoming very difficult. For eg., 2.5TB of videos/images for consumption are reality today. But the data pipelines to consume big data to algorithms have not been evolved enough to be usable, except by a few top internet companies.

In this panel, the panelists will present their point of view on pressing next challenges concerning Big Data Heterogeneity. The moderator will leverage a diverse set of experiences and viewpoints of the panel members and encourage them to share their controversial points of view and provocative positions.

**Moderator:** Vijay Raghavan, Alfred & Helen Lamson Professor, University of Louisiana at Lafayette, USA

#### **Panel Members:**

Jiawei Han, Abel Bliss Professor, UIUC, USA

Patrick Valduriez, Senior Researcher, LeanXcale, Inria, France

Ramanathan Guha, Founder and Lead, DataCommons.org, Google, USA

Sharad Mehrotra, Professor, UC Irvine, USA

# Posters

Poster ID	Accept Posters
P202	Lili Xiang, Alice Mello, and Ryan Ackerman, <i>Launching a User-Generated Content Campaign to Promote PUMA's Social Contributions</i>
P204	zhengwu sun, <i>Big data analysis of social networking</i>
P205	Sumedh Yadav, Gautam Kumar, and Shivam Kumar, <i>A graph construction study for graph-based semi-supervised learning: Case study on unstructured text data</i>
P207	Murtadha Kareem and Oliver Faust, <i>Establishing the safety of a smart heart health monitoring service through validation</i>
P208	Georg Heiler and Allan Hanbury, <i>Comparing Implementation Variants Of Distributed Spatial Join on Spark</i>
P209	Burak Cetin, Alina Lazar, Jinoh Kim, Alex Sim, and Kesheng Wu, <i>Federated Wireless Network Intrusion Detection</i>
P211	Edouard Ngor SARR, Ousmane SALL, Mamadou BOUSSO, Rabiyaou DIOUF, Babiga BIRREGAH, and Sény Ndiaye MBAYE, <i>Automatic Categorization of Press Articles through Learning: The Case of Senegalese Online Press</i>
P212	Rabiyaou DIOUF, Edouard Ngor SARR, Ousmane SALL, Mamadou BOUSSO, Babiga BIRREGAH, and Sény Ndiaye MBAYE, <i>Web Scraping: State-of-the-Art and Areas of Application</i>
P214	Ji Cheng, Guimu Guo, Da Yan, Xiaotian Hao, and Wilfred Ng, <i>EasyRain: A User-Friendly Platform for Comparing Precipitation Nowcasting Models</i>
P215	Guimu Guo, Jalal Majed Khalil, Da Yan, and Virginia Sisiopiku, <i>Realistic Transport Simulation with Open Data</i>
P216	Alex Aizman, Gavin Maltby, and Thomas Breuel, <i>High Performance I/O For Large Scale Deep Learning</i>
P217	Manoj M, <i>Federated Query processing for Big data in Data Science</i>
P218	Kenta Yamada and Hayato Yamana, <i>Effectiveness of Usability &amp; Performance Features for Web Credibility Evaluation</i>
P219	Monika Ray, Banafsheh Sadeghi, Dominique Ritley, and Patrick S. Romano, <i>Impact of Mandated Public Reporting in California on 30-Day readmission following CABG surgery: A Health policy analysis</i>
P220	Ranjeet Devarakonda, Giri Prakash, Kavya Guntupally, and Jitendra Kumar, <i>Big Federal Data Centers Implementing FAIR Data Principles: ARM Data Center Example</i>
P221	Eric Bax and John Donald, <i>Sharp Frequency Bounds for Sample-Based Queries</i>
P222	Koji Iwanuma, Takumi Nishina, and Yoshitaka Yamamoto, <i>Accelerating an On-Line Approximation Mining for Large Closed Itemsets</i>
P223	Byron Gao and Frank Medjo, <i>Statistical Correction of Average Customer Ratings for Product Ranking</i>
P224	Timothy John Ebido, KiChul Park, and Kyungho Jeon, <i>Targot: Improving Dataset Upload Time to Object Storage using Client-Server Cooperation</i>
P225	Itay Lieder, Meirav Segal, Eran Avidan, Asaf Cohen, and Tom Hope, <i>Learning a faceted customer segmentation for discovering new business opportunities at Intel</i>
P226	Yifan Ding, Daheng Wang, Tim Weninger, and Meng Jiang, <i>Preserving Composition and Crystal Structures of Chemical Compounds in Atomic Embedding</i>
P227	simon woo, Siho Han, and Yaup Kim, <i>Classifying Genuine Face images from Disguised Face Images</i>
P228	Emebo Onyeka, Vaibhav Anu, and Aparna S. Varde, <i>Identifying Implicit Requirements in SRS Big Data</i>
P229	Byron Gao and Gayathri Karupakula Jagadeesh Kumar, <i>CoRank: Simultaneously Ranking Publication Venues and Researchers</i>
P230	Nikodimos Provatatos, Ioannis Konstantinou, and Nectarios Koziris, <i>Towards Faster Distributed Deep Learning Using Data Hashing Techniques</i>
P231	Evdokia Kassela, Ioannis Konstantinou, and Nectarios Koziris, <i>Towards a Multi-engine Query Optimizer for Complex SQL Queries on Big Data</i>
P232	Hsing Bung Chen, Zhi Qiao, and Song Fu, <i>Applying SDN based data network on HPC Big Data Computing – Design, Implementation, and Evaluation</i>
P233	Zhichao Wang, Shashank Singh, and Arnold Pereira, <i>Large Scale Time Series Analysis for Infrastructure Reliability</i>
P234	Darlan Arruda and Nazim H. Madhavji, <i>QualiBD: A Tool for Modelling Quality Requirements for Big Data Applications</i>
P235	Yun Ning Pek and Kwan Hui Lim, <i>Identifying and Understanding Business Trends using Topic Models with Word Embedding</i>

P236	Frederique van Leeuwen, <i>Utilizing Multivariate Time Series for Semantic Segmentation</i>
P237	Evdokia Kassela, Nikodimos Provatas, Asterios Tsiourvas, Ioannis Konstantinou, and Nectarios Koziris, <i>BigOptiBase: Big Data Analytics for Base Station Energy Consumption Optimization</i>
P239	Choudur Lakshminarayan, Thiagarajan Ramakrishnan, Awany AlOmari, Khaled Bouaziz, Faraz Ahmad, Sri Raghavan, and Prama Agarwal, <i>Model Management and Handwritten Character Recognition: An Enterprise Solution</i>
P240	Youness Arjoune, Sai Peri, Niroop Sugunraj, Debanjan Sadhukhan, Michael Nord, Gautham Krishnamoorthy, David Flynn, and Prakash Ranganathan, <i>Thermal Imagery Based Instance Segmentation for Energy Audit Applications in Buildings</i>
P241	Crisrael Lucero and Phuong Nguyen, <i>A Modern Approach to Big Provenance</i>
P242	Emad Felemban, Abdulaziz Fatani, and Faizan Ur Rehman, <i>An Optimized Scheduling Process for a Large Crowd to Perform Spatio-temporal Movements Safely during Pilgrimage</i>
P243	Mehrad Eslami, Yicheng Tu, Hadi Charkhgard, Zichen Xu, and Jiacheng Liu, <i>PsiDB: A Framework for Batched Query Processing and Optimization</i>
P244	Parag Paul and Manas Sharma, <i>Q95-squared, a singular metric to define histogram quality for a database system</i>
P245	Sylvie Koziel, <i>Application of big data analytics to support power networks and their transition towards smart grids</i>
P247	Masafumi Oyamada, <i>Extracting Feature Engineering Knowledge from Data Science Notebooks</i>
P248	EunJeong Hwang and Yong-Hyuk Kim, <i>Interdependency between the Stock Market and Financial News</i>
P249	Shuan Chen and Hyun Uk Kim, <i>Designing Novel Functional Peptides by Manipulating a Temperature in the Softmax Function Coupled with Variational Autoencoder</i>
P250	Yong Jin, Masahiko Tomoishi, and Satoshi Matsuura, <i>Detection of Hijacked Authoritative DNS Servers by Name Resolution Traffic Classification</i>
P251	Sukhwan Jung and Wan Chul Yoon, <i>Citation-based Author Contribution Measure for Byline-Independency</i>
P252	Abbas Mazloui and Rajiv Gupta, <i>Enabling Faster Convergence in Distributed Irregular Graph Processing</i>
P253	Li Gao and Weikai He, <i>On Understanding Biosonar Deformations Using Deep Learning-Based Video Interpolation</i>
P254	Li Gao, Hongjie Jiang, Kaiming Fu, and Weikai He, <i>On Understanding Degradation Kinetics of Pharmaceutical Gelatin Matrices for Precision Medicine: A Deep Learning Approach</i>
P255	Sung-Wook Choi, Chang-Seob Song, and Chae-Soo Kim, <i>PARTS CATALOG OBJECT RECOGNITION TECHNOLOGY FOR EFFICIENT DRAWING DISTRIBUTION MANAGEMENT</i>
P256	Chae-Soo Kim, Jun-Yeong Gwon, and Seong-Beom Son, <i>A Study on Solution Oriented Smart Factory Diagnostic System for SME</i>
P257	Rosangela Casolare, Fabio Martinelli, Francesco Mercaldo, and Antonella Santone, <i>A Model Checking based Proposal for Mobile Colluding Attack Detection</i>
P258	Fabio Di Tommaso, Michele Guerra, Fabio Martinelli, Francesco Mercaldo, Massimo Piedimonte, Giovanni Rosa, and Antonella Santone, <i>User Authentication through Keystroke Dynamics by means of Model Checking: A Proposal</i>
P259	Giovanni Capobianco, Umberto Di Giacomo, Francesco Mercaldo, Antonella Santone, and Tommaso Di Tusa, <i>A Methodology for Real-Time Data Verification exploiting Deep Learning and Model Checking</i>
P260	Essa Alhazmi and Nazim Choudhury, <i>Sign Prediction in Online Games</i>
P261	Roshan Bharath Das, Marc X. Makkes, Alex Uta, Lin Wang, and Henri Bal, <i>A Programming Framework for Heterogeneous Stream Analytics</i>
P262	Joseph Cauthen, Marco Mercado, Tejas Gandre, Meetkumar Patel, and Mohammad Husain, <i>An Authentication System using Neurological Responses to Music</i>
P263	Cenru Liu, <i>A Hierarchical Classification Model for Solar Flare Prediction</i>
P264	Cristina Menghini, Aris Anagnostopoulos, and Eli Upfal, <i>Wikipedia Polarization and Its Effects on Navigation Paths</i>
P265	Kouichi Nagatani, Chisa Takano, and Masaki Aida, <i>Spectral Analysis of User Interests for Experimental Verification of the Oscillation Model for OSNs</i>
P266	Eugene Choy, Winston Ho, Xiaohang Li, Ragini Verma, Li Sim, and Kyong Jin Shim, <i>Happy Toilet: A Social Analytics Approach to the Study of Public Toilet Cleanliness</i>
P267	Rujing Yao, Linlin Hou, Yingchun Ye, Ou Wu, Ji Zhang, and Jian Wu, <i>Method and Dataset Mining in Scientific Papers</i>
P268	Eugene Choy, Gladys Ng, Martius Lim, and Kyong Jin Shim, <i>Listen, Nudge, Empower: A Mobile Gratitude Journal Application</i>
P269	Anthony James, Frank Gonzales, Matt Kedis, Micheal Balestrieri, Navid Ahmadiyeh, Araya Gebeyehu, Richard Lam, and Karina Rowe, <i>Integrated Grid Analytics - DER Situational Awareness &amp; Susceptibility</i>
P270	Jane Seah and Kyong Jin Shim, <i>Plugin to a Healthier Life: A Web Browser Plugin for Mental Health Monitoring</i>

P271	Cheng Lee, Joel-David Wong, Zi Lim, Belinda Tho, Sean Kwek, and Kyong Jin Shim, <i>How Does Fake News Spread: Raising Awareness &amp; Educating the Public with a Simulation Tool</i>
P272	Maruthi Prithivirajan and Kyong Jin Shim, <i>An IoT-Driven Smart Cafe Solution for Human Traffic Management</i>
P273	marc capelo, Karan Aggarwal, and Pranjul Yadav, <i>Combining Text and Image data for Product Recommendability Modeling</i>
P274	Daisuke Maeda and Sudhanshu Gaur, <i>Remaining Useful Life Prediction of Industrial Consumables Using Wideband Vibration Signals</i>
P275	Weikai He and Li Gao, <i>A Statistical Causal Inference Method for Exploring Ultrasonics and Topological Deformations in Biological Systems</i>
P276	Ibrahim Almubark, Lin-Ching Chang, Thanh Nguyen, Raymond Turner, and Xiong Jiang, <i>Early Detection of Alzheimer's Disease Using Patient Neuropsychological and Cognitive Data and Machine Learning Techniques</i>
P278	Charles Grumer, Jonathan Peck, Femi Olumofin, Anderson Nascimento, and Martine De Cock, <i>Hardening DGA Classifiers Utilizing IVAP</i>
P279	Kashyap Venkatesh, Bashar Barmada, Veronica Liesaputra, and Guillermo Ramirez-Prado, <i>Robust Features for Activities Recognition</i>
P280	Tao Li, Minsoo Choi, Kaiming Fu, Li Gao, and Lei Lin, <i>Music Sequence Prediction with Mixture Hidden Markov Models</i>
P281	Warren Landis, Sangwhan Cha, and Majid Shaalan, <i>On Optimization of Stock Market Prediction Methods</i>
P282	Haodi Qi, Hanyu Jiang, Wende Bu, Chengzi Zhang, and Kyong Jin Shim, <i>Tracking Political Events in Social Media: A Case Study of Hong Kong Protests</i>
P283	Xuejiao Tang, Jiong Qiu, Wenbin Zhang, Ibrahim Toure, Mingli Zhang, Enza Messina, Xueping Xie, Xuebing Wang, and Sheng Yu, <i>The Internet of Responsibilities - Connecting Human Responsibilities using Big Data and Blockchain</i>
P284	Kihoon Jang, Junwhan Kim, and Byunggu Yu, <i>Vector-based Churn Prediction using Neural Networks in Mobile Games</i>
P285	Wei Wang, Changyang Feng, and Wei Quan, <i>Data-driven control of a class of discrete-time linear complex dynamical networks</i>
P286	Amir Behrouzi-Far and Emina Soljanin, <i>Data Replication for Reducing Computing Time in Distributed Systems with Straggles</i>
P287	Amir Behrouzi-Far and Emina Soljanin, <i>Scheduling in the Presence of Data Intensive Compute Jobs</i>
P288	Guillermo Ramirez-Prado, Bashar Barmada, and Veronica Liesaputra, <i>On non-intrusive prediction of activities and behavior</i>
P289	Abrar Rahman, Yonathan Weiner, Hailey Lynn Swanson, Rebecca Slepian, Anusheh Abdullah, and Marvin Slepian, <i>A Big-Data Approach to Defining Breathing Signatures for Identifying Respiratory Disease</i>
P290	Hani Ramadhan, Yoga Yustiawan, and Joonho Kwon, <i>Extracting valid indoor semantic trajectories using movement constraints</i>
P291	Cheng-Yu Chen and Shun-Wen Hsiao, <i>IoT Malware Dynamic Analysis Profiling System and Family Behavior Analysis</i>
P292	Mohsen Karimi, Ali Jahanshahi, Abbas Mazloumi, and Hadi Zamani Sabzi, <i>Border Gateway Protocol Anomaly Detection Using Neural Network</i>
P293	Lok Ting Lam and Shun-Wen Hsiao, <i>AI-Based Online P2P Lending Risk Assessment On Social Network Data With Missing Value</i>
P294	Junkun Peng, Pin Ni, Jiayi Zhu, Zhenjin Dai, Yuming Li, Gangmin Li, and Xuming Bai, <i>Automatic Generation of Electronic Medical Record Based on GPT2 Model</i>
P295	Kun Yuan Sung and Shun Wen Hsiao, <i>Mitigating DDoS with PoW and Game Theory</i>
P296	Jiayi Zhu, Pin Ni, Yuming Li, Junkun Peng, Zhenjin Dai, Gangmin Li, and Xuming Bai, <i>A Word2vec Based on Chinese Biomedical Domain Knowledge</i>
P297	Xingang Wang, Renshi Yu, Tailian Liu, Wenqing Li, and Xiaoling Sun, <i>Map Adjustment As a Base for Privacy Assurance in Semantic Spatial Trajectories Release</i>
P298	Alireza Abdoli, Amy Murillo, Alec Gerry, and Eamonn Keogh, <i>Time Series Classification: Lessons Learned in the (Literal) Field while Studying Chicken Behavior</i>
P299	Kazuma Kusu and Kenji Hatano, <i>Recurrent Path Index for Efficient Graph Traversal</i>
P300	Abrar Rahman, Ari Mitra, Fuad Rahman, and Marvin Slepian, <i>Smart EHR – A Big-Data Approach to Automated Collection and Processing of Multi-Modal Health Signals in a Doctor-patient Encounter</i>
P301	Lilian Ngweta, Karan Bhanot, Ariane Maharaj, Ian Bogle, and Thilanka Munasinghe, <i>Identifying the Relationship Between Precipitation and Zika Outbreaks in Argentina</i>
P302	Najib Ishaq, George Student, and Noah Daniels, <i>Clustered Hierarchical Entropy-Scaling Search of Astronomical and Biological Data</i>
P303	Joseph Sebastian and Thilanka Munasinghe, <i>Airline Award Miles Redemption</i>
P304	SHILPA BALAN, TANVI GAWADE, and AAKANKSHA TASGAONKAR, <i>A Machine Learning Approach for Prediction of Length of Stay for the Kid's Inpatient Database</i>

P305	Anh Thai, Vy Bui, Laura Reyes, and Lin-Ching Chang, <i>Using Deep Convolutional Neural Network for Mouse Brain Segmentation in DT-MRI</i>
P306	Thilanka Munasinghe, Evan Patton, and Oshani Seneviratne, <i>IoT Application Development Using MIT App Inventor to Collect and Analyze Sensor Data</i>
P308	Teja Reddy, Kush Arya, and Zachary Dodds, <i>The Limits to the Efficiency of Machine Learning</i>
P312	Jiwan Lee, Bonghee Hong, Chumsoo Kim, and Woo Chan, <i>Development of a Continuous Complex Event Processing Platform for Real-Time Tactical Moving Objects</i>
P313	Jhih-Jie Chen, Yi-Dong Wu, Yu-Chuan Tai, Ching-Yu Yang, Hai-Lun Tu, and Jyun-Sheng Chang, <i>Extracting Grammatical Error Corrections from Wikipedia Revision History</i>
P314	Hai-Lun Tu, Pei-Chen Ho, Jason S. Chang, and Li-Guang Chen, <i>Word Sense Disambiguation Using Wikipedia Link Graph</i>
P315	Pin Ni, Yuming Li, Jiayi Zhu, Junkun Peng, Zhenjin Dai, Gangmin Li, and Xuming Bai, <i>Disease Diagnosis Prediction of EMR Based on BiGRU-Att-CapsNetwork Model</i>
P316	John Pomerat, Aviv Segev, and Rituparna Datta, <i>On Neural Network Activation Functions and Optimizers in Relation to Polynomial Regression</i>
P317	Yuming Li, Pin Ni, Junkun Peng, Jiayi Zhu, Zhenjin Dai, Gangmin Li, and Xuming Bai, <i>A Joint Model of Clinical Domain Classification and Slot Filling Based on RCNN and BiGRU-CRF</i>
N212	Choudur Lakshminarayan, Thiagarajan Ramakrishnan, Awny AlOmari, Khaled Bouaziz, Faraz Ahmad, Sri Raghavan, and Prama Agarwal, <i>Enterprise-wide Machine Learning using Teradata Vantage: An Integrated Analytics Platform</i>
N225	Ruihong Huang, Zhiwei Chen, Zhicheng Liu, Shaoxu Song, and Jianmin Wang, <i>TsOutlier: Explaining Outliers with Uniform Profiles over IoT Data</i>
N228	Neil Brockett, Catriona Clarke, Michele Berlingerio, and Sourav Dutta, <i>A System for Analysis and Remediation of Attrition</i>
N231	Aleksandr Suleikin, Peter Panfilov, and Natalya Bakhtadze, <i>Industrial track: Architecting railway KPIs data processing with Big Data technologies</i>
N237	David Huber, Nigel Stepp, Aruna Jammalamadaka, Tiffany Kim, Sam Johnson, Dana Warmesley, and Tsai-Ching Lu, <i>MATRICES: A System for Human-Machine Hybrid Forecasting of Geopolitical Events</i>
N240	Su Won Bae, Aravind Ravi, Kajanana Sangaralingam, Nisha Verma, Anindya Datta, and Varun Chugh, <i>Suspicious Location Detection Using Trajectory Analysis &amp; Location Backfilling - A Scalable Approach</i>
N241	Yu Cheng and Shuo Yang, <i>An Interactive Online Audience Extension System</i>

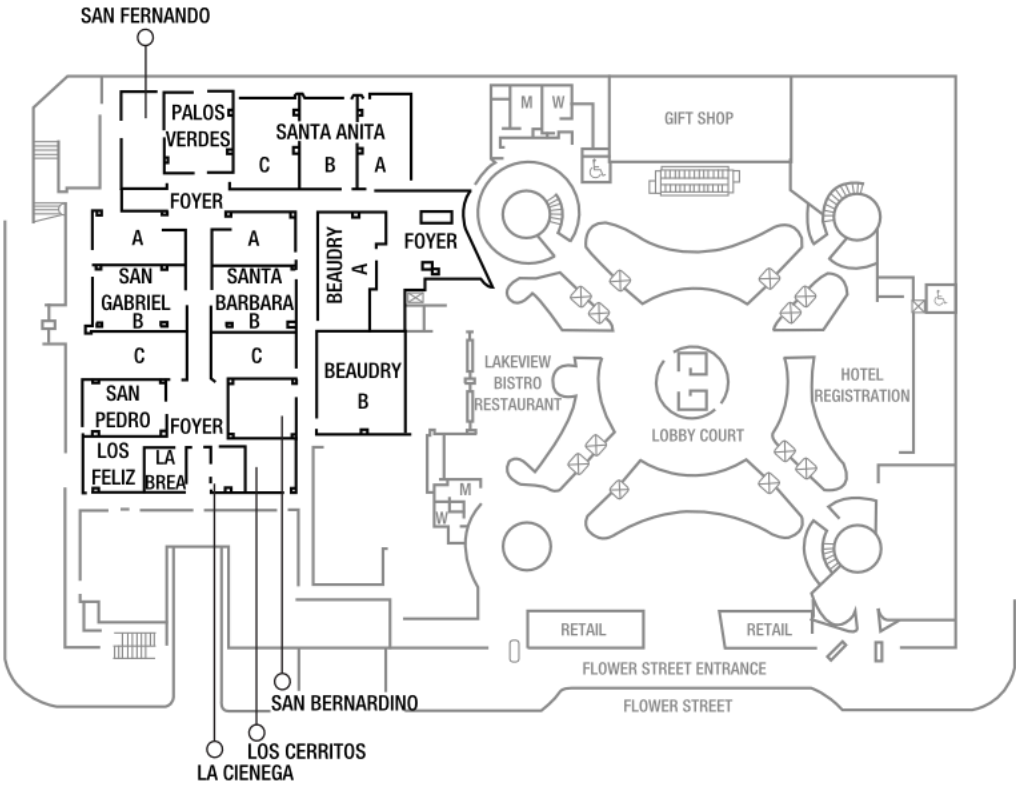
# **2019 Big Data Conference Wifi Access**

**Connect to network: BIG DATA  
Password: bigdata2019**

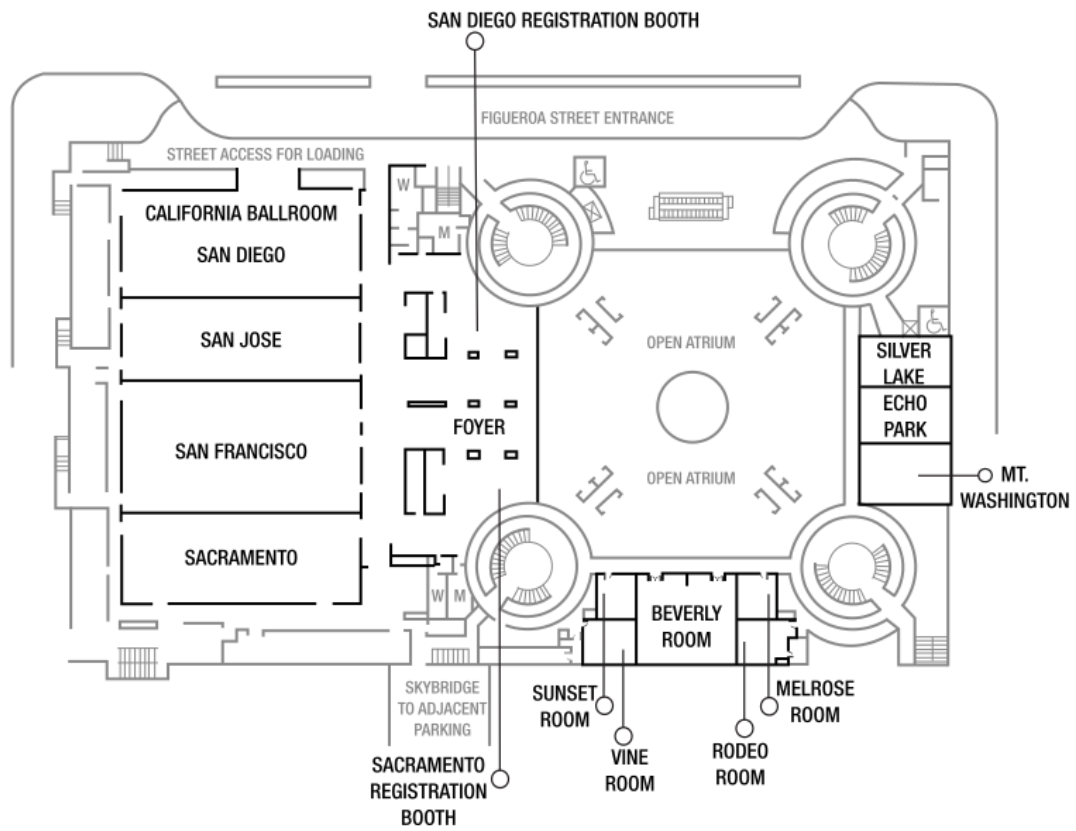


# Westin Hotel Floor Plan

## LOBBY LEVEL - FUNCTION ROOMS



## LEVEL 2 - CALIFORNIA BALLROOM



## IEEE BIGDATA 2020

December 10-13, 2020, Atlanta, GA, 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 Security & Privacy

- a. PGA/CGRA/GPU accelerator for Big Data applications
- b. Operating system support and runtimes for hardware accelerators
- c. Programming models and platforms for accelerators
- d. Domain-specific and heterogeneous architectures
- e. Novel system organizations and designs
- f. Computation in memory /storage/network
- g. Persistent, non-volatile and emerging memory for big data
- h. Operating system support for high-performance network architectures

### 7. 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 the use of Big Data in industry. We accept full papers (up to 10 pages) and extended abstracts (2-4 pages).

