



Original Image by Richo Fan (CC BY 2.0)

# **45TH**

## **INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES**

**August 26-30, 2019**

**Los Angeles, California**



# 45<sup>TH</sup> INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES



Logo developed by Hamed Alimohammadzadeh

## CONFERENCE PROGRAM

**AUGUST 26-30, 2019**

**LOS ANGELES, CALIFORNIA**

**<https://vldb.org/2019>**



# TABLE OF CONTENTS

General Chair’s Welcome Message . . . . .	4
Program Committee Chairs’ Welcome Message . . . . .	5
Conference Officers . . . . .	7
Student Volunteers . . . . .	21
2019 VLDB Endowment Awards . . . . .	22
2019 VLDB Conference Awards . . . . .	23
DBCares . . . . .	25
Reception . . . . .	26
Banquet . . . . .	27
At the Conference . . . . .	28
Conference Venue . . . . .	30
Camp VLDB . . . . .	31
Internet Access . . . . .	31
Conference Map . . . . .	33
Our Sponsors . . . . .	38
Program Overview . . . . .	41
Detailed Program . . . . .	47
Detailed Workshop Program . . . . .	78
Demo Sessions: Group A and B . . . . .	104
Posters of VLDBJ . . . . .	109
VLDB 2020 . . . . .	111

# GENERAL CHAIR'S WELCOME MESSAGE

The annual VLDB conference is a premier international forum for database researchers, vendors, practitioners, application developers, and users. VLDB 2019 is held at Los Angeles, California. The Los Angeles basin is home to premier academic institutions, think tanks, major studios and movie production facilities, and Silicon Beach with many innovative startups. This year's VLDB has a separate but equally prestigious research and industry tracks for the papers presented at the conference. Eleven workshops on timely topics ranging from Polysores to Blockchain and Internet-of-Things are co-located with VLDB 2019, providing tremendous opportunities for technological and entrepreneurial innovation.

In addition to a strong technical program, the generous support of our industrial sponsors has enabled us to offer unique social events including a Welcome reception on Monday August 26th, a reception at the California Science Center with the Space Shuttle Endeavor on Tuesday August 27th, and a conference Banquet on Wednesday August 28th. The sponsor support has also enabled us to offer competitive registration rates for both regular and student attendees. This year, we are offering a day pass to accommodate those traveling from San Diego, the Bay Area, Seattle, and Vancouver to experience VLDB for one conference day. We thank our industrial sponsors. We are also grateful to the National Science Foundation for providing student support (and the under-represented groups) to attend VLDB 2019.

VLDB 2019 would not be possible if it were not for dedicated service of numerous people from prestigious institutions. Their names are listed throughout this booklet and I make no attempt to repeat their names here. I am grateful to each and everyone for their valuable service. I also acknowledge and thank participants.

I hope you enjoy VLDB 2019 and its outstanding program.

General Conference Chair  
**Shahram Ghandeharizadeh, USC**

# PROGRAM COMMITTEE CHAIRS' WELCOME MESSAGE

International Conference on Very Large Data Bases (VLDB) is a premier annual international forum for data management and data analytics researchers, industry practitioners, application developers, and end users. As the foundations of sophisticated algorithms and scalable solutions, data management plays a prominent role in machine learning and artificial intelligence research. The annual conference often consists of a mix of keynote talks, peer reviewed research and industrial presentations, panel discussions, tutorials, demonstrations, invited industrial talks, and workshops.

VLDB conference provides a unique forum for the research papers accepted to the **Proceedings of the VLDB Endowment (PVLDB)** journal to be presented in a conference setting. VLDB 2019 includes papers from PVLDB volume 12, as well as 7 roll over papers from PVLDB volume 11. The Review Board for Volume 12 of PVLDB consists of 166 expert researchers, and reviewing is coordinated by 17 Associate Editors. Each volume of PVLDB provides 12 monthly submissions, with a quick 4 weeks of first round reviews, followed by two weeks of active discussions led by the Associate Editors. Revision papers have up to 3 months to submit their revised submissions.

This year's VLDB conference will take place in Los Angeles, California, from August 26 to August 30, 2019. VLDB 2019 covers many aspects of data management and analytics, including data integration, cloud databases, distributed transactions, query processing and optimization, crowdsourcing, graph analytics, scalable machine learning, and distributed systems.

After 15 months of careful and dedicated hard work by the Review Board, Associate Editors, and Editors in Chief of PVLDB volume 12, we have selected 128 papers to be presented at VLDB 2019. These submissions come from 52 different countries and territories from all over the world, with 332 submissions from U.S.A, 151 submissions from P. R. China, and 54 submissions from Germany as the most popular countries. PVLDB Volume 12 received 677 research submissions this year. Out of 677 submissions includes 587 regular research papers, 8 vision papers, 31 innovation systems and applications papers, and 51 experimental and analysis papers. The conference program also includes 7 "roll-over" papers from PVLDB Volume 11, for a total of 135 research papers. In addition to research presentations, the full conference program also includes 22 peer-reviewed industrial papers selected out of 72 submissions, 48 research demonstrations selected out of 127 submissions, 7 tutorials selected out of 15 submissions, and a panel on inter-disciplinary research opportunities between data management and AI/ML. The conference features three plenary keynote presentations. Finally, the conference includes two poster sessions with all the industrial and research papers, as well as 15 posters of VLDB journal papers. On Monday August 26, and Friday August 30, the conference is accompanied by 11

workshops on topics of particular interest to the data management community.

The VLDB 2019 technical program reflects the efforts of hundreds of members of the research community who have prepared papers, demonstrations and presentations for the conference. We would like to thank to all authors who have submitted contributions. Especially, we would like to thank the Review Board and the Associate editors for the countless hours they spent to ensure high quality reviews, and fair and professional discussions. It is their dedication and diligent work that makes VLDB a premier forum for database researchers. Finally, we also would like to thank the industrial chairs Wolfgang Lehner, Pat Helland, and Beng Chin Ooi, demonstration chairs Nesime Tatbul and Alin Deutsch, tutorial chairs Xin Luna Dong and Amr El Abbadi, and the panel chairs Sang Cha and Tamer Özsu for putting together an impressive program. We are looking forward to an outstanding technical program, and we hope you enjoy the conference.

**Fatma Özcan**, IBM Research - Almaden

**Lei Chen**, Hong Kong University of Science and Technology

PVLDB Volume 12 Editors in Chief

VLDB 2019 Program Committee Chairs



# CONFERENCE OFFICERS

***General Conference Chairs***

Shahram Ghandeharizadeh, USC

***VLDB 2019 Program Committee Chairs***

***PVLDB Volume 12 Editors in Chief***

Lei Chen, HKUST

Fatma Özcan, IBM Research - Almaden

***VLDB Endowment Representative***

Michael Carey, UC - Irvine

***Proceedings Chairs***

Abdul Quamar, IBM Research - Almaden

Yongxin Tong, Beihang University

***Tutorial Chairs***

Amr El Abbadi, UC - Santa Barbara

Xin Luna Dong, Amazon

***Industrial Chairs***

Beng Chin Ooi, NUS - Singapore

Pat Helland, Salesforce

Wolfgang Lehner, TU Dresden

***Demo Chairs***

Alin Deutsch, UC - San Diego

Nesime Tatbul, Intel Labs and MIT

***Panel Chairs***

Sang Kyun Cha, National Seoul University

M. Tamer Özsu, University of Waterloo

***PhD Workshop Chairs***

Ilaria Bartolini, University of Bologna

Feifei Li, Alibaba - USA

***Workshop Chairs***

Sharad Mehrotra, UC - Irvine

Yuanyuan Tian, IBM Research - Almaden

***Treasury Chair***

K. Selçuk Candan, ASU

***Liaison for Industrial Sponsors***

Chen Li, UC - Irvine

***Sponsorship Committee Chairs***

Anand Deshpande, Persistent Systems

Xiaoyong Du, Renmin University

Volker Markl, TU Berlin

Renee Miller, Northeastern

***Publicity Committee Chairs***

Sumita Barahmand, Microsoft

Jason Yap, Google

***Web/Information Chair***

Mehran Barahmand, Amazon

***Local Arrangements Chairs***

Tyson Condie, Microsoft

Beth Trushkowsky, HMC

***Registration Chairs***

Ahmed Eldawy, UC - Riverside

Faisal Nawab, UC - Santa Cruz

***System Chair***

Mohammad Sadoghi, UC - Davis

***Booklet Chair***

Mariam Salloum, UC - Riverside

***Volunteer Coordinators***

Shobeir Fakhraei, USC

Xiang Ren, USC

# **ASSOCIATE EDITORS OF PVLDB VOLUME 12**

Azza Abouzied, NYU Abu Dhabi, UAE  
A. Selçuk Candan, Arizona State University, USA  
Surajit Chaudhuri, Microsoft Research, USA  
Amol Desphande, University of Maryland, USA  
Johann-Christoph Freytag, Humboldt-Universität zu Berlin, Germany  
Rainer Gemulla, University of Mannheim, Germany  
Nick Koudas, University of Toronto, Canada  
Georgia Koutrika, Athena Research Center, Greece  
Yun Yao Li, IBM Research - Almaden, USA  
Alexandra Meliou, University of Massachusetts, USA  
Arnab Nandi, Ohio State University, USA  
M. Tamer Özsu, University of Waterloo, Canada  
Themis Palpanas, University of Paris, France  
Neoklis (Alkis) Polyzotis, Google Research, USA  
Kyuseok Shim, Seoul National University, South Korea  
Xiaokui Xiao, National University of Singapore, Singapore  
Meihui Zhang, Beijing Institute of Technology, China

# REVIEW BOARD OF PVLDB VOLUME 12

Ashraf Abounaga, QCRI  
Divyakant Agrawal, UCSB  
Ismail Sengor Altingovde, METU - Turkey  
Ilkay Altintas, San Diego Supercomputing Center  
Yael Amsterdamer, Bar-Ilan University  
Nicolas Anciaux, INRIA  
Pelin Angin, METU - Turkey  
Arvind Arasu, Microsoft  
Nikolaus Augsten, University of Salzburg  
Ilaria Bartolini, University of Bologna  
Philip Bernstein, Microsoft Research  
Arnab Bhattacharya, IIT Kanpur  
Spyros Blanas, Ohio State University  
Matthias Boehm, Graz University of Technology  
Michael Böhlen, University of Zurich  
Angela Bonifati, University of Lyon - France  
Philippe Bonnet, ITU - Copenhagen  
Sebastian Breß, DFKI - TU Berlin  
Huiping Cao, NMSU  
Sang Kyun Cha, Seoul National University  
Chee-Yong Chan, NUS - Singapore  
Badrish Chandramouli, Microsoft Research  
Yi Chen, NJIT  
Hong Cheng, Chinese University of Hong Kong  
James Cheng, Chinese University of Hong Kong  
Fei Chiang, McMaster University  
Byron Choi, Hong Kong Baptist University  
Xu Chu, Georgia Tech  
Bobbie Cochrane, IBM  
Tyson Condie, Microsoft  
Bin Cui, Peking University - China  
Carlo Curino, Microsoft  
Khuzaima Daudjee, University of Waterloo  
Bolin Ding, Alibaba  
Jens Dittrich, Saarland University - Germany  
Xin Luna Dong, Amazon  
Curtis Dyreson, Utah State University  
Mohamed Eltabakh, WPI  
Alan Fekete, University of Sydney  
Hakan Ferhatosmanoglu, University of Warwick  
Avrilia Floratou, Microsoft

Ada Wai-Chee Fu, Chinese University of Hong Kong  
Johann Gamper, Free University of Bozen-Bolzano - Italy  
Jinyang Gao, NUS - Singapore  
Yunjun Gao, Zhejiang University  
Tingjian Ge, University of Massachusetts - Lowell  
Buğra Gedik, Bilkent University - Turkey  
Floris Geerts, University of Antwerp  
Boris Glavic, Illinois Institute of Technology  
Zhiguo Gong, University of Macau  
Goetz Graefe, Google  
Hakan Hacigumus, Google  
Wook-Shin Han, Postech - Korea  
Oktie Hassanzadeh, IBM Research - Yorktown  
Michael Hay, Colgate University  
Melanie Herschel, University of Stuttgart - Germany  
Katja Hose, Aalborg University  
Zi Huang, University of Queensland  
Seung-Won Hwang, Yonsei University  
Yoshiharu Ishikawa, Nagoya University  
H. V. Jagadish, University of Michigan  
Eser Kandogan, IBM Research - Almaden  
Murat Kantarcioglu, University of Texas at Dallas  
Pinar Karagoz, METU - Turkey  
Manos Karpathiotakis, Facebook  
Arijit Khan, Nanyang Technological University  
Daniel Kifer, Penn State University  
Donald Kossmann, Microsoft Research  
Arun Kumar, UC - San Diego  
Chuan Lei, IBM Research - Almaden  
Viktor Leis, Friedrich Schiller University Jena  
Chen Li, UC - Irvine  
Chengkai Li, UT Arlington  
Feifei Li, Alibaba - USA  
Guoliang Li, Tsinghua University  
Min Li, IBM Research - Almaden  
Xiang Lian, Kent State University  
Xuemin Lin, University of New South Wales  
Hua Lu, Aalborg University  
Ashwin Machanavajjhala, Duke University  
Stefan Manegold, CWI Amsterdam  
Ioana Manolescu, INRIA  
Essam M. Mansour, Concordia University – Canada  
Michael Mathioudakis, University of Helsinki  
Umar Farooq Minhas, Microsoft Research  
Mohamed F. Mokbel, Qatar Computing Research Institute  
Mirella Moro, Universidade Federal de Minas Gerais  
Davide Mottin, Aarhus University  
Barzan Mozafari, University of Michigan  
Parth Nagarkar, NMSU  
Vivek Narasayya, Microsoft Research  
Mario Nascimento, University of Alberta - Canada  
Azade Nazi, Microsoft Research  
Raymond Ng, UBC  
Beng Chin Ooi, NUS - Singapore

Vincent Oria, NJIT  
George Papadakis, University of Athens  
Dimitris Papadias, HKUST  
Olga Papaemmanouil, Brandeis University  
Paolo Papotti, EURECOM - France  
Stefano Paraboschi, Università degli Studi di Bergamo  
Jignesh Patel, University of Wisconsin - Madison  
Andrew Pavlo, CMU  
Lucian Popa, IBM Research - Almaden  
Danica Porobic, Oracle  
Fabio Porto, LNCC - Brazil  
Abdul Quamar, IBM Research - Almaden  
Vijayshankar Raman, Google  
Maya Ramanath, IIT Delhi  
Berthold Reinwald, IBM Research - Almaden  
Matthias Renz, George Mason University  
Mirek Riedewald, Northeastern University  
Mary Roth, IBM Research - Almaden  
Senjuti Basu Roy, New Jersey Institute of Technology  
Sudip Roy, Google  
Florin Rusu, UC Merced  
Diptikalyan Saha, IBM Research - India  
Semih Salihoglu, University of Waterloo  
Karthik Sankaranarayanan, IBM Research - India  
Maria Luisa Sapino, University of Torino - Italy  
Ahmet Erdem Sariyuce, University at Buffalo  
Mohamed Sarwat, ASU  
Yücel Saygin, Sabanci University - Turkey  
Shuo Shang, KAUST - Saudi Arabia  
Yingxia Shao, BUPT  
Heng Tao Shen, UESTC - China  
Yanyan Shen, Shanghai Jiao Tong University  
Alkis Simitsis, Microfocus  
Ambuj Singh, UCSB  
Shaoxu Song, Tsinghua University  
Kostas Stefanidis, Tampere University  
Letizia Tanca, Politecnico di Milano  
Nan Tang, QCRI  
Egemen Tanin, University of Melbourne - Australia  
Jens Teubner, TU Dortmund  
Martin Theobald, University of Luxembourg  
Yuanyuan Tian, IBM Research - Almaden  
Hanghang Tong, ASU  
Yongxin Tong, Beihang University  
Pinar Tozun, ITU - Copenhagen  
Immanuel Trummer, Cornell University  
Yannis Velegrakis, Utrecht University  
Stratis Viglas, University of Edinburgh  
Sai Wu, Zhejiang University  
Yinghui Wu, Washington State University  
Yingjun Wu, IBM Research - Almaden  
Li Xiong, Emory University  
Jianliang Xu, HKBU  
Jun Yang, Duke University

Xiaochun Yang, Northeastern University  
Junjie Yao, East China Normal University  
Hongzhi Yin, The University of Queensland  
Cong Yu, Google  
Ye Yuan, Beijing Institute of Technology  
Demetrios Zeinalipour-Yazti, University of Cyprus  
Steffen Zeuch, DFKI - TU Berlin  
Wenjie Zhang, UNSW  
Ying Zhang, UTS  
Kai Zheng, UESTC - China  
Xiangmin Zhou, RMIT  
Xiaofang Zhou, University of Queensland  
Kostas Zoumpatianos, Harvard University

# INDUSTRIAL TRACK PC MEMBERS

Divy Agrawal, University of California, Santa Barbara  
Pedro Bizarro, Feedzai  
Alexander Boehm, SAP SE  
Ben Busjaeger, Salesforce  
Mark Callaghan, Facebook  
Michael Carey, UC Irvine  
Angelo Ciarlini, McKinsey  
Anh Dinh, SUTD  
Thomas Fanghaenel, Salesforce  
Avrilia Floratou, Microsoft  
Anil Goel, SAP  
Alon Halevy, Recruit  
Stavros Harizopoulos, Facebook  
Lars Hofhansl, Salesforce  
Theodore Johnson, AT&T Labs - Research  
Yannis Katsis, IBM Research - Almaden  
Martin Kersten, CWI  
Kihong Kim, SAP Labs - Korea  
Dustin Lange, Amazon Research  
Feifei Li, Alibaba - USA  
Yongzhu (Eric) Li, Futurewei  
Yunyao Li, IBM Research - Almaden  
Qian Lin, National University of Singapore  
Wei Liu, Tencent  
Wei Lu, Renmin University of China  
Rui Mao, Shenzhen University  
Jamie Martin, Salesforce  
Neoklis Polyzotis, Google  
Danica Porobic, Oracle  
Berthold Reinwald, IBM Research - Almaden  
Amit Rustagi, Western Digital  
Benjamin Schlegel, Oracle Labs  
Pinar Tozun, ITU  
Dimitris Tsirogiannis, Amazon, Inc.  
Hannes Voigt, Neo4j  
Guoren Wang, Beijing Institute of Technology  
Wei Wang, NUS  
Reynold Xin, Databricks  
Yang Yang, University of Electronic Science and Technology of China  
Kai Zeng, Alibaba - China  
Feida Zhu, Singapore Management University  
Arthur Zwiegincew, Oculus/Facebook



# DEMO TRACK PC MEMBERS

Nesreen Ahmed, Intel Labs - USA  
Lyublena Antova, Datometry - USA  
Manos Athanassoulis, Boston University - USA  
Cagri Balkesen, Oracle Labs - Switzerland  
Leilani Battle, University of Maryland - USA  
Carsten Binnig, TU Darmstadt - Germany  
Angela Bonifati, University of Lyon - France  
Renata Borovica-Gajic, University of Melbourne - Australia  
Vanessa Braganholo, Fluminense Federal University - Brazil  
Lei Cao, MIT - USA  
Alvin Cheung, UC Berkeley - USA  
Laura Chiticariu, IBM Data & AI - USA  
Sudipto Das, Amazon Web Services - USA  
Cagatay Demiralp, Megagon Labs - USA  
Dong Deng, Rutgers University - USA  
Jens Dittrich, Saarland University - Germany  
Jiang Du, Facebook - USA  
Iman Elghandour, IT University of Copenhagen - Denmark  
David Eyers, University of Otago - New Zealand  
Peter Fischer, University of Augsburg - Germany  
Vijay Gadepally, MIT Lincoln Laboratory - USA  
Avigdor Gal, Technion - Israel  
Wolfgang Gatterbauer, Northeastern University - USA  
Buğra Gedik, Bilkent University - Turkey  
Ioana Giurgiu, IBM Research - Switzerland  
Boris Glavic, Illinois Tech - USA  
Michael Grossniklaus, University of Konstanz - Germany  
Annika Hinze, University of Waikato - New Zealand  
Katja Hose, Aalborg University - Denmark  
Alekh Jindal, Microsoft - USA  
Pinar Karagoz, METU - Turkey  
Flip Korn, Google Research - USA  
Georgia Koutrika, Athena Research Center - Greece  
Justin Levandoski, Amazon - USA  
Feifei Li, Alibaba - USA  
Eric Lo, CUHK - Hong Kong  
Qiong Luo, HKUST - Hong Kong  
Alex Moga, ABB Research - Switzerland  
Daniela Nicklas, University of Bamberg - Germany  
Silvia Nittel, University of Maine - USA  
Themis Palpanas, Paris Descartes University - France  
John Paparrizos, University of Chicago - USA  
Aditya Parameswaran, UIUC - USA

Peter Pietzuch, Imperial College London - UK  
Karthik Ramasamy, Streamlio - USA  
Alexander Rasin, DePaul University - USA  
Semih Salihoglu, University of Waterloo - Canada  
Nadathur Satish, Facebook - USA  
Marco Serafini, UMass Amherst - USA  
Kyumars Sheykh Esmaili, TomTom - Belgium  
Ben Sowell, Amazon - USA  
Rebecca Taft, Cockroach Labs - USA  
Nan Tang, QCRI - Qatar  
Richard Tibbetts, Tableau - USA  
Srikanta Tirthapura, Iowa State University - USA  
Kristin Tufte, Portland State University - USA  
Marcos Antonio Vaz Salles, University of Copenhagen - Denmark  
Matthias Weidlich, Humboldt University of Berlin - Germany  
Xiangyao Yu, MIT - USA

# PHD WORKSHOP PC MEMBERS

Sourav S. Bhowmick, Nanyang Technological University  
K. Selçuk Candan, ASU  
Reynold Cheng, University of Hong Kong  
Guoliang Li, Tsinghua University  
Faisal Nawab, UC Santa Cruz  
Vincent Oria, NJIT  
M. Tamer Özsu, University of Waterloo  
Dimitris Papadias, HKUST  
Marco Patella, University of Bologna  
Jeff Phillips, University of Utah  
Maria Luisa Sapino, University of Torino  
Mohamed Sarwat, ASU  
Divesh Srivastava, AT&T Labs-Research  
Yufei Tao, Chinese University of Hong Kong  
Yingjun Wu, IBM Research - Almaden  
Ke Yi, HKUST

# EXTERNAL REVIEWERS

Michael Abebe, University of Waterloo  
Cuney Akcora, University of Texas at Dallas  
Ahmed Al-Baghdadi, Kent State University  
Mohammad Javad Amiri, UC - Santa Barbara  
Imrul Anindya, University of Texas at Dallas  
Aref Asvadishirehjini, University of Texas at Dallas  
Davide Azzalini, Politecnico di Milano  
Fabio Azzalini, Politecnico di Milano  
Qiushi Bai, UC - Irvine  
Peter Bailis, Stanford University  
Fuat Basik, Bilkent University  
Srikanta Bedathur, IIT-Delhi  
Lu Can, The Chinese University of Hong Kong  
Yang Cao, Kyoto University  
Pingfu Chao, The University of Queensland  
Chris Conlan, University of Warwick  
Camila Costa, University of Alberta  
Teddy Cunningham, University of Warwick  
Xinyan Dai, The Chinese University of Hong Kong  
Bonaventura Del Monte, DFKI GmbH  
Bailu Ding, Microsoft Research  
Anh Dinh, National University of Singapore  
Anshuman Dutt, Microsoft Research  
Jingzhi Fang, Beihang University  
Yixiang Fang, The University of New South Wales  
Daniel Ford, IBM  
Junhao Gan, University of Queensland  
Brad Glasbergen, University of Waterloo  
Sergey Gorbunov, University of Waterloo  
Bingsheng He, National University of Singapore  
Pedro Holanda, CWI  
Wen Hua, University of Queensland  
Jianfeng Jia, Google  
Su Jiao, The Chinese University of Hong Kong  
Fengmei Jin, Renmin University of China  
Tatiana Jin, The Chinese University of Hong Kong  
Alex Jones, UC - Santa Barbara  
Taewoo Kim, UC - Irvine  
Kostas Kollias, Google Research  
Ios Kotsogiannis, Duke University  
Georgia Koutrika, Athena Research Institute  
Avinash Kumar, UC - Irvine  
Longbin Lai, The University of New South Wales

Guoliang Li, Tsinghua University  
Jingjing Li, UESTC  
Lei Li, University of Queensland  
Qian Lin, National University of Singapore  
Boge Liu, The University of New South Wales  
Zhi Liu, The Chinese University of Hong Kong  
David Lomet, Microsoft Research  
Mark Ma, Emory University  
Sujaya Maiya, UC - Santa Barbara  
Sourav Medya, UC - Santa Barbara  
Li Meiyong, National University of Singapore  
Mehmet Ercan Nergiz, Sabanci University  
Sarah Oppold, University of Stuttgart  
Mustafa Ozdayi, University of Texas at Dallas  
Niketan Pansare, Apple  
Marco Patella, DISI, University of Bologna  
Todd Porter, Microsoft  
Navneet Potti, UW-Madison  
Niranjan Rai, Kent State University  
Nisarg Raval, Duke University  
Weilong Ren, Kent State University  
Sami Saeef, University of Texas at Arlington  
Sadeem Saleh, UC - Irvine  
Felix Martin Schuhknecht, Saarland University  
Prithviraj Sen, IBM Research - Almaden  
Rathijit Sen, Microsoft  
Neha Sengupta, IIT-Delhi  
Huasong Shan, JD.com American Technologies Corporation  
Ankur Sharma, Saarland University  
Xianghai Sheng, UC - Merced  
Arlei Silva, UC - Santa Barbara  
Panos Simatis, HKUST  
Stavros Sintos, Duke University  
Diego Tome, CWI  
Jonas Traub, TU Berlin  
Dimitris Tsakalidis, HKUST  
Dimitris Tsaras, HKUST  
Daniel Garcia Ulloa, Keysight Technologies  
Shivaram Venkataraman, Microsoft Research  
Vinu Venugopal, University of Luxemburg  
Stratis Viglas, Google  
Hanchen Wang, The University of New South Wales  
Sen Wang, The University of Queensland  
Sheng Wang, National University of Singapore  
TaiNing Wang, National University of Singapore  
Wei Wang, National University of Singapore  
Wei Wang, The University of New South Wales  
Zuozhi Wang, UC - Irvine  
Kevin Wellenzohn, University of Zurich  
Dong Wen, The University of Technology Sydney  
Lingfei Wu, IBM Research - Yorktown  
Yidi Wu, The Chinese University of Hong Kong  
Yingjun Wu, IBM Research - Almaden  
Cheng Xu, Hong Kong Baptist University

Xiao Yan, The Chinese University of Hong Kong  
Zhengyi Yang, The University of New South Wales  
Kai Yao, HKUST  
Wei Ye, UC - Santa Barbara  
Fan Zhang, The University of New South Wales  
Jian Zhang, The Chinese University of Hong Kong  
Zuyu Zhang, UW-Madison  
Jianwen Zhao, The Chinese University of Hong Kong  
Weijie Zhao, UC, Merced  
Yunjian Zhao, The Chinese University of Hong Kong

# STUDENT VOLUNTEERS

Tsatsral Amarbayasgalan, Chungbuk National University  
Mohammad Javad Amiri, UC Santa Barbara  
Asoke Datta, UC Merced  
Karima Echihabi, Mohammed V University  
Suyash Gupta, UC Davis  
Shohedul Hasan, University of Texas Arlington  
Sona Hasani, University of Texas Arlington  
Haoyu Huang, USC  
Shiva Jahangiri, UC Irvine  
Sajjad Rahnama, UC Davis  
Paulus Shituna, University of Namibia  
Abhishek Singh, UC Santa Cruz

# 2019 VLDB ENDOWMENT AWARDS

## EARLY CAREER RESEARCH CONTRIBUTION AWARD

**Aditya Parameswaran**, for developing tools for large-scale data exploration, targeting non-programmers.

## WOMEN IN DATABASE RESEARCH AWARD

**Wang-Chiew Tan**, for contributions to data provenance and integration spanning theory and practice.

## TEST OF TIME AWARD

**HadoopDB: An Architectural Hybrid of MapReduce and DBMS Technologies for Analytical Workloads. PVLDB 2(1): 922-933 (2009)**

*Authors: Azza Abouzeid, Kamil Bajda-Pawlikowski, Daniel Abadi, Avi Silberschatz, and Alexander Rasin*

**PNUTS: Yahoo!'s hosted data serving platform. PVLDB 1(2): 1277-1288 (2008)**

*Authors: Brian F. Cooper, Raghuram Ramakrishnan, Utkarsh Srivastava, Adam Silberstein, Philip Bohannon, Hans-Arno Jacobsen, Nick Puz, Daniel Weaver, and Ramana Yerneni*

A decade ago was a time of deep architectural change in database systems, with changes occurring on both the OLTP side and the OLAP side. PNUTS and HadoopDB exemplify that transformation. Both had industrial impact, and they influenced significant follow-on research, as indicated by in excess of 2,000 citations.



# 2019 VLDB CONFERENCE AWARDS

## BEST PAPER AWARD:

### **Fine-Grained, Secure and Efficient Data Provenance for Blockchain**

*Authors: Pingcheng Ruan (National University of Singapore), Gang Chen (Zhejiang University), Anh Dinh (National University of Singapore), Qian Lin (National University of Singapore), Beng Chin Ooi (NUS), Meihui Zhang (Beijing Institute of Technology)*

## HONORABLE MENTION

### **Declarative Recursive Computation on an RDBMS, or, Why You Should Use a Database For Distributed Machine Learning**

*Authors: Dimitrije Jankov (Rice University), Shangyu Luo (Rice University), Binhang Yuan (Rice University), Zhuhua Cai (Rice University), Jia Zou (Rice university), Chris Jermaine (Rice University), Zekai Gao (Rice University)*

## DISTINGUISHED REVIEWERS

Arun Kumar, UCSD  
Boris Glavic, Illinois Institute of Technology  
Buğra Gedik, Bilkent University  
Byron Choi, Hong Kong Baptist University  
Danica Porobic, Oracle  
Florin Rusu, UC - Merced  
Floris Geerst, Univ of Antwerp  
Jianliang Xu, HKBU  
Jun Yang, Duke University  
Li Xiong, Emory University  
Matthias Boehm, Graz University of Technology  
Melanie Herschel, Universität Stuttgart  
Michael Hay, Colgate University  
Paolo Papotti, Eurocom  
Philip Berstein, MSR  
Pinar Tozun, ITU  
Semih Salihoglu, University of Waterloo  
Shaoxu Song, TsingHua University  
Stefano Paraboschi, Università degli Studi di Bergamo  
Tingjian Ge, University of Massachusetts, Lowell  
Umar Farooq Minhas, MSR  
Xu Chu, Georgia Institute of Technology  
Yanyan Shen, Shanghai Jiaotong University

Yinghui Wu, Washington State University  
Yongxin Tong, Beihang University  
Yuan Ye, Northeastern University  
Yuanyuan Tian, IBM Research - Almaden

# DBCARES

VLDB 2019 adheres to the ACM Policy Against Harassment. We strive to ensure that VLDB 2019 is carried out in an inclusive and diverse environment with zero tolerance for discrimination, harassment, or any other form of misconduct.

Examples of unacceptable behavior at VLDB are:

- **Abuse:** Any action directed at an individual that (a) interferes substantially with that person's participation; or (b) causes that person to fear for his/her personal safety. This includes threats, intimidation, bullying, stalking, or other types of abuse.
- **Discriminatory Harassment:** Any conduct that discriminates or denigrates an individual on the basis of race, ethnicity, religion, citizenship, nationality, age, sexual or gender identity, disability, or any other characteristic protected by law in the location where the ACM activity takes place.
- **Sexual Harassment:** Unwelcome sexual advances, requests for sexual favors, or other verbal/physical conduct of a sexual nature.

Any participant who experiences unacceptable behavior may contact any current member of the VLDB Executive Committee (<https://www.vldb.org/trustees.html>) or DBCares ([http://www.vldb.org/vldb\\_cares.html](http://www.vldb.org/vldb_cares.html)). Please be assured that if you approach us, your concerns will be kept in strict confidence, and we will consult with you before any actions are taken.

# RECEPTION

Join us for an evening of scientific exhibits at the California Science Center.

The California Science Center, The Oschin  
Space Shuttle Endeavour Pavilion  
27 August 2019, 7 p.m. to 10 p.m. PST  
700 Exposition Park Dr.  
Los Angeles, CA 90037  
+1 (323) 724-3623



## RECEPTION DETAILS

- The reception will be located in the Oschin Space Shuttle Endeavour Pavilion, home to the historic space shuttle Endeavor. The Endeavor is the fifth orbiter space shuttle built for NASA's space shuttle program. The spacecraft went on 25 missions from 1992-2011; it is now on display at the California Science Center.
- See the Endeavor up close as you enjoy hors d'oeuvres and beverages (alcoholic and non-alcoholic). The pavilion includes videos and displays where you can learn more about this amazing spacecraft. Note: those with big appetites may want to grab dinner before the reception.

## EXHIBITS

- During the event you will also have access to the eight interactive exhibits that are part of the Ecosystems Gallery. Each exhibit explores a different ecological principle. For example, in the Forest Zone you can check out a kelp forest through a tunnel at the bottom of an 188,000-gallon tank; in the Extreme Zone you can learn about rocky shore organisms such as sea stars (and even touch a few!). Enjoy live animals and hands-on experiences throughout in the Ecosystems Gallery.

## TRANSPORTATION

- VLDB 2019 will provide bus service to and from the reception. Loading for buses will begin at 17:45 at the hotel's Figueroa Street entrance, located on Level 2. Return trips leaving the California Science Center will begin at 21:00 and drop at the pickup location. The last bus leaves the California Science Center at 22:30.

# BANQUET

Join us for an evening of food, live music, and dancing at the VLDB 2019 banquet dinner!

## BANQUET DETAILS

- The banquet dinner will be held in the Pasadena Room, located on the level below the lobby at the Westin Bonaventure. This buffet dinner is a wonderful opportunity to sit and relax with colleagues over a glass of wine.

## ENTERTAINMENT

- The beginning of the evening includes a live performance by "The Off-By-One Quartet" band consisting of (alphabetically) Jaime Aclander (AT&T, guitar), Mike Carey (UC Irvine & Couchbase, bass), Joe Hellerstein (UC Berkeley & Trifacta, trumpet), 7:30-8:00 pm.
- Afterwards, get ready to dance your heart out! Grab your colleagues and head to the dance floor for continued entertainment provided by DJ Taha.

# AT THE CONFERENCE

## QR CODE

- Each badge contains a QR code to enable participants and sponsors to connect with one another. Each QR code contains the participant's name, affiliation, and e-mail address.

## PARKING

- **On-site parking**, fee: 15 USD hourly, 49 USD daily
- **Valet parking**, fee: 49 USD daily
- Parking for vehicles up to 6 feet tall. For oversized vehicles, contact the hotel for parking alternatives

## PUBLIC TRANSPORT

- **Subway Station**: 0.3 miles (500 m) south from the hotel, located at 7th and Flower Street; this is the best way to get to the Santa Monica Beach.
- **Train Station**: Union Station, is 1.6 miles (2.6 km) northeast from the hotel.

## DRUGSTORES

- **CVS**: 812 S Grand Ave, Los Angeles, CA 90017, +1 (213) 627-7925. Pharmacy open until 9PM M-F, and 6PM Sat-Sun. Store open until 10PM daily for over-the-counter drugs.
- **Rite-aid**: 600 W 7th St, Los Angeles, CA 90017, +1 (213) 896-0083. Pharmacy open until 9PM M-F, and 6PM Sat-Sun. Store open until midnight daily for over-the-counter drugs.
- **Walgreens**: 617 W 7th St, Los Angeles, CA 90017, +1 (213) 694-2880. Pharmacy open until 9PM M-F, and 6PM Sat-Sun. Store open until 10PM daily for over-the-counter drugs.

## MEDICAL EMERGENCY

- Immediately call or ask someone to call 911 (only three digits)
- **Dignity Health California Hospital**: 1401 S Grand Ave, Los Angeles, CA 90015, +1 (213) 770-0276.

## NEARBY DINING OPTIONS

- **LA Prime** (Steakhouse): Savor mouthwatering steak and fresh seafood as you take in spectacular 360-degree views of LA from LA Prime, the award-winning steakhouse on the 35th floor of downtown's The Westin Bonaventure Hotel & Suites, Los Angeles. Open for dinner. Dress code: Casual. +1 (213) 612-4743; ext: 4743.

- **BonaVista Lounge** (American): Sip classic cocktails made with an innovative twist at the dazzling BonaVista Lounge. Tempting libations, light restaurant entrees and small plates are on the menu, as are terrific views of Los Angeles from the 34th floor of this revolving hotel venue. Open for lunch and dinner. Dress code: Casual. +1 (213) 612-4743; ext: 4743
- **Lakeview Bistro** (American): Enjoy breakfasts and lunch favorites at Lakeview Bistro. This popular downtown LA restaurant offers healthy American classics, and an impressive list of fine wines. Our Westin's Super Foods program will support your commitment to becoming a healthier you. Open for breakfast and lunch. Dress code: Casual. +1 (213) 612-4897; ext: 4897
- **Lobby Court** (American): Stop in for your favorite beverage at Lobby Court. Grab a Starbucks® coffee to start your day or sip cocktails and sample appetizers after a long day of meetings or exploring Los Angeles. Open for breakfast, lunch, and dinner. Dress code: Casual. +1 (213) 624-1000; ext: 4897
- **Border Grill, Downtown LA** (Mexican): Border Grill is a contemporary Mexican restaurant in Downtown LA, offering fresh, handmade margaritas and bold flavors of modern Mexican fare. Open for lunch and dinner. Dress code: Casual.
- **Drago Centro** (Italian): Drago Centre is an upscale Italian dining experience, combining classic Italian recipes and techniques with a more modern panache. Open for lunch and dinner. Dress code: Smart Casual.
- **George's Greek Grill** (Mediterranean): George's Greek Grill specializes in providing fresh, tasty sandwiches, super-wraps and other Mediterranean favorites. Open for lunch and dinner. Dress code: Casual.
- **71Above Restaurant and Skylounge** (American): Enjoy beautiful panoramic views of LA and high-end American cuisine while dining in a lap of luxury. Open for lunch and dinner. Dress code: Smart Casual.

## SUPERMARKETS

- **Whole Foods:** Supermarket specializing in organic and natural foods. 788 S Grand Ave, Los Angeles, CA 90017, +1 (213) 873-4745.
- **Smart & Final Extra!:** Warehouse-like supermarket chain for produce, meats & packaged foods, plus discounts on bulk items. 845 S Figueroa St #100, Los Angeles, CA 90017, +1 (213) 629-0039.

# CONFERENCE VENUE

The Westin Bonaventure Hotel & Suites  
404 S Figueroa St.  
Los Angeles, CA 90071  
+1 (213) 624-1000



Los Angeles is more than just a city—it's an adventure. Beaches, whale-watching, and hiking. Celebrities and colorful locals. Temperate weather. Captivating attractions, museums, and festivals. The Westin Bonaventure Hotel & Suites, Los Angeles places you in the center of all the excitement. Our downtown Los Angeles hotel spans an entire city block in the heart of the bustling business district, near the city's best sights and sounds.

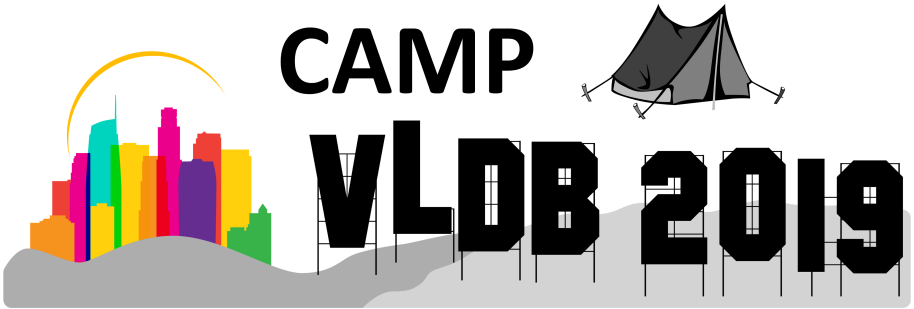
At 367-foot (112 m) and 35-stories tall, the Westin Bonaventure Hotel and Suites was designed by architect John C. Portman, Jr., and constructed between 1974 and 1976. It is the largest hotel in the city and has been featured in many movies and television series over the years including: *Interstellar*, *Strange Days*, *Buck Rogers in the 25th Century* (as part of the city of New Chicago), *Blue Thunder*, *It's a Living*, *L.A. Law*, *Breathless*, *Matlock*, *This Is Spinal Tap*, *Hit the Booty Do*, *Nick of Time*, *Midnight Madness*, *Showtime*, *Hard to Kill*, *The Lincoln Lawyer*, *Chuck*, *Xanadu*, *Moby Dick*, *The Fantastic Journey* and was destroyed (via special effects) in *Escape from LA*, *Epicenter*, and *San Andreas*. Experience one of America's legendary metropolises while staying in one of our comfortable, chic hotel rooms.

Its located at the epicenter of the city's financial district, less than two blocks from the Union Bank and US Bank towers, with a skyway connecting us to downtown's World Trade Center. Leisure is equally within reach: we're the closest upscale hotel to LA's famed Dodger Stadium. Also right in our vicinity is the Figueroa Corridor/University Park area, home to University of Southern California (USC). This area encompasses the Los Angeles Memorial Coliseum, Exposition Park and Shrine Auditorium, the National History Museum, California African American Museum, and the California Science Center & IMAX theater. Stop by the city's newly-built Cathedral of Our Lady of the Angels, less than a mile from the hotel.

As the entertainment capital of the world, LA is one of the country's most glamorous cities—only one component of which is its thriving film industry. Among LA's foremost addresses, we are a short stroll to the Music Center, the Museum of Contemporary Art, and Frank Gehry's Walt Disney Concert Hall. We're also only eight short blocks to the Los Angeles Convention Center, LA Live, and the celebrated Staples Center, and just 20 minutes from Los Angeles International Airport (LAX).



# CAMP VLDB



**Monday, August 26**  
8:00 AM – 6:00 PM

**Tuesday, August 27**  
8:00 AM – 6:00 PM

**Wednesday, August 28**  
8:00 AM – 6:00 PM

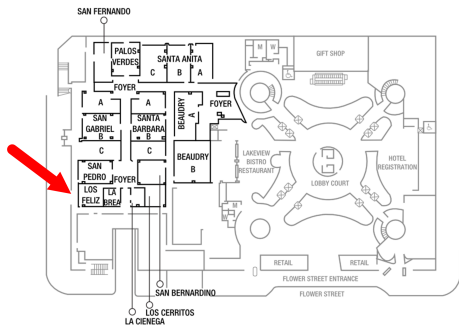
**Thursday, August 29**  
8:00 AM – 6:00 PM

**Friday, August 30**  
8:00 AM – 6:00 PM

For children ages 6 months to 5 years

### Location: Los Feliz Room

LOBBY LEVEL - FUNCTION ROOMS



Camp VLDB 2019 is sponsored by



# Microsoft

# INTERNET ACCESS

Internet access is provided for participants  
at the conference floor.

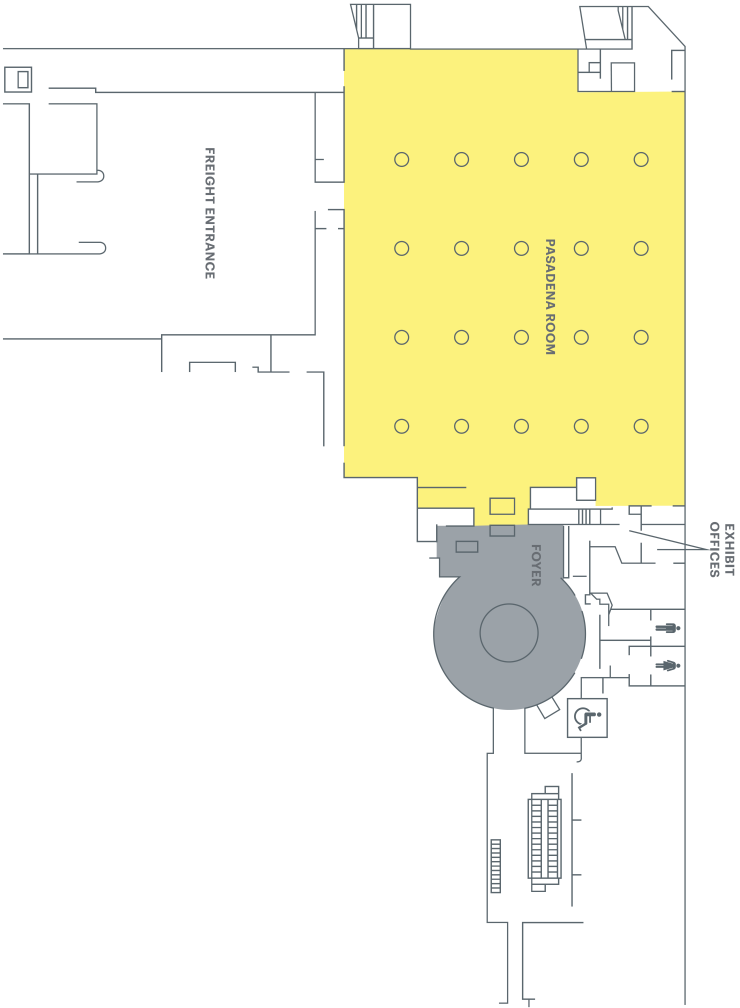
Follow the instructions below:

VLDB 2019 Open Wi-fi  
Network Name (SSID): **VLDB**  
Password: **VLDB2019**

Valid during the event schedule  
from Monday, August 26  
to Friday, August 30.

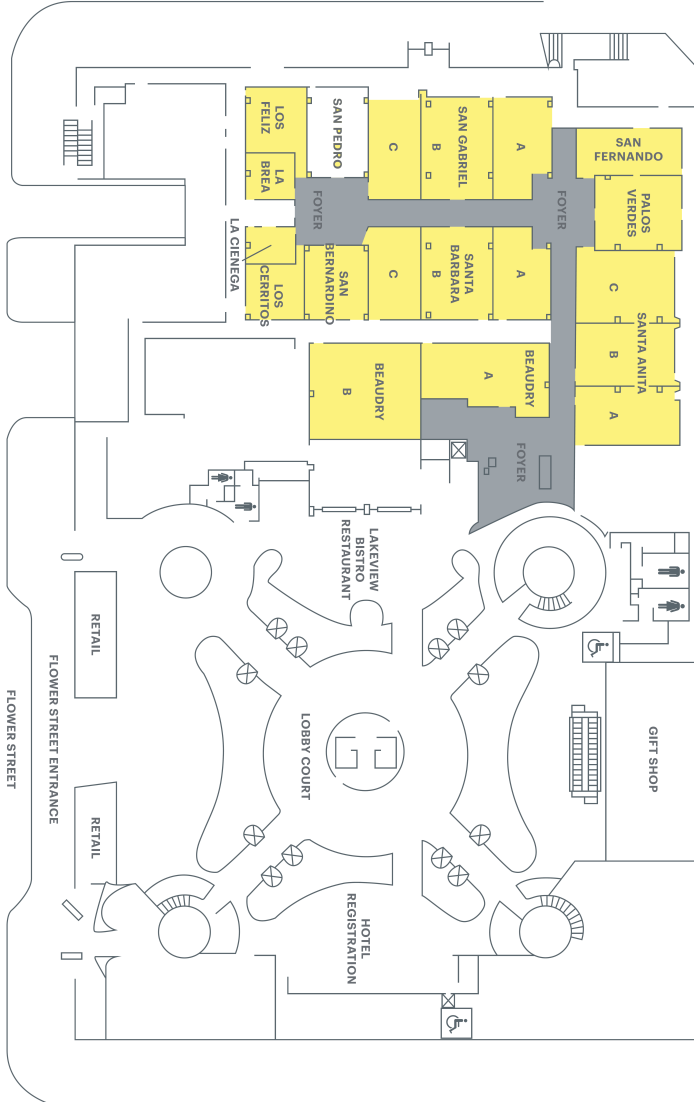
# CONFERENCE MAP

ONE LEVEL BELOW LOBBY



One Level Below Lobby: Pasadena Room Exhibition Hall

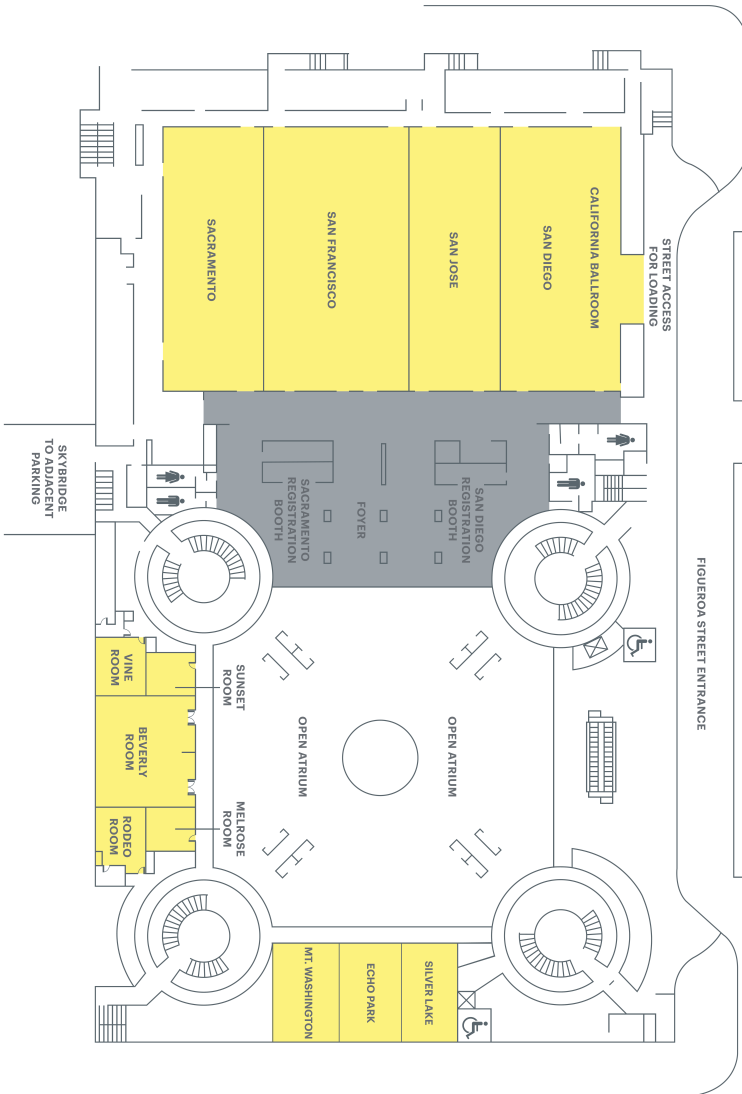
**LOBBY LEVEL**



Lobby Level Function Rooms

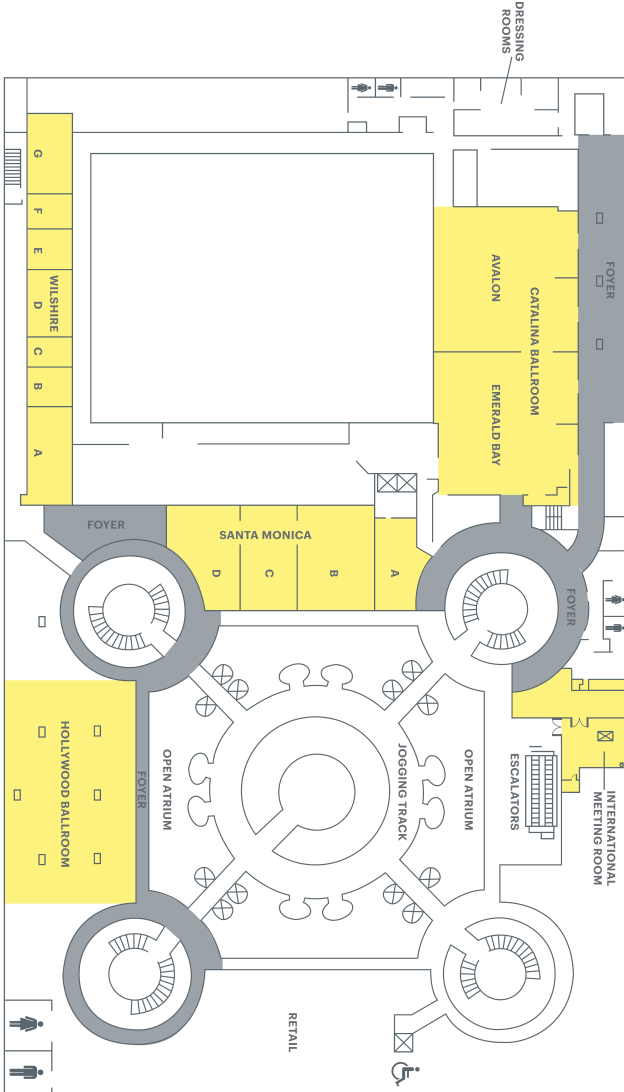
LEVEL 2

Level 2 Function Rooms



LEVEL 3

Level 3 Function Rooms



# OUR SPONSORS

## DIAMOND SPONSORS





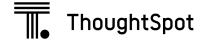
PLATINUM SPONSORS



GOLD SPONSORS



SILVER SPONSORS



BRONZE SPONSORS



EXHIBITORS



# **PROGRAM OVERVIEW**

Day 1: Monday (August 26, 2019)					
Time	Room: San Jose	Room: San Diego	Room: Emerald Bay	Room: Avalon	Room: Santa Anita
9:00-10:30	AIDB	BIRTE	ADMS	TPCTC	PHD Workshop
10:30-11:00	Coffee Break - 30 min (Room: California Foyer)				
11:00-12:30	AIDB	BIRTE	ADMS	TPCTC	PHD Workshop
12:30-14:00	Lunch Break - 90 min (Room: Plaza Deck)				
14:00-15:30	AIDB	BIRTE	ADMS	TPCTC	PHD Workshop
15:30-16:00	Coffee Break - 30 min (Room: California Foyer)				
16:00-18:30	AIDB	BIRTE	ADMS	TPCTC	PHD Workshop
19:00-22:00	Welcome Reception (Room: Plaza Deck)				

Day 2: Tuesday (August 27, 2019)							
Time	Room: Santa Anita	Room: San Jose	Room: San Diego	Room: Emerald Bay	Room: Hollywood Ballroom	Room: Avalon	Room: Sacramento / San Francisco
8:00-8:30							
8:30-9:00	Morning Coffee (Room: California Foyer)						
9:00-10:30	Conference Inauguration (Room: Sacramento / San Francisco)						
10:30-11:00	Keynote 1: Tamer M. Özsu – Graph Processing - A Panoramic View and Some Open Problems (Room: Sacramento/San Francisco)						
11:00-12:30	Coffee Break - 30 min (Room: California Foyer)						
	<b>Research Session 1:</b> Data-driven Machine Learning	<b>Research Session 2:</b> Distributed Graphs	<b>Research Session 3:</b> Transaction Processing I	<b>Research Session 4:</b> Query Processing I	<b>Demo Session:</b> Group A	<b>Tutorial 1:</b> Data Lake Management: Challenges and Opportunities	<b>Industry Session 1:</b> Deep Learning and Graph DBs
12:30-14:00	Lunch Break - 90 mins (Room: Pasadena)						
14:00-15:30	Keynote 2: Tova Milo - Getting Rid of Data (Room: Sacramento / San Francisco )						
15:30-16:00	Coffee Break - 30 min (Room: California Foyer)						
16:00-17:30	<b>Research Session 5:</b> Cloud Databases	<b>Research Session 6:</b> Subgraphs and Communities	<b>Research Session 7:</b> Similarity Search and Data Extraction	<b>Research Session 8:</b> Hardware and Machine Learning	<b>Demo Session:</b> Group B	<b>Tutorial 2:</b> Combating Fake News: A Data Management and Mining Perspective	<b>Industry Invited Talks 1:</b> Data Standing Tall in the Clouds
17:45-18:30	Departure for the Reception, California Science Center						
19:00-22:00	VLDB Conference Reception at the California Science Center						

Day 3: Wednesday (August 28, 2019)

Time	Room:	Room:	Room:	Room:	Room:	Room:
	Santa Anita	San Jose	San Diego	Emerald Bay	Hollywood Ballroom	Sacramento / San Francisco

Morning Coffee (Room: California Foyer)

8:30-9:00 DBCares - Chaudhuri, VLDB Endowment - Markl, VLDB 2020 - Ishikawa (Room: Sacramento / San Francisco)

9:00-10:30 Keynote 3: Eric Iverson - Awesome Data Use Cases from Hollywood: How data is changing everything about the entertainment experience (Room: Sacramento / San Francisco )

10:30-11:00 Coffee Break - 30 min (Room: California Foyer)

11:00-12:30	<b>Research Session 9:</b> Distributed Systems I	<b>Research Session 10:</b> Graph Analytics I	<b>Research Session 11:</b> Stream Processing and Analysis	<b>Research Session 12:</b> Database Engines for New Hardware	<b>Demo Session:</b> Group A	<b>Tutorial 3:</b> Speedup Your Analytics: Automatic Parameter Tuning for Databases and Big Data Systems	<b>Industry Session 2:</b> High Performance DBs: Both General Purpose and Specialized
-------------	--	---	--	---	------------------------------	--	---

12:30-14:00 Lunch Break - 90 min (on your own, see list of restaurants on page 28)

14:00-15:30 Panel: Opportunities for data-management research in the era of horizontal AI/ML (Room: Sacramento / San Francisco )

15:30-16:00 Coffee Break - 30 min (Room: California Foyer)

16:00-17:30	<b>Research Session 13:</b> Query Optimization	<b>Research Session 14:</b> Data Integration and Crowdsourcing	<b>Research Session 15:</b> New Architectures and Caching	<b>Research Session 16:</b> Data Exploration	<b>Demo Session:</b> Group B	<b>Tutorial 4:</b> The Ever Evolving Online Labor Market: Overview, Challenges and Opportunities	<b>Invited Industry Talks 2:</b> Data through the Looking Glass: Performance, In-Mem., & AI
-------------	--	--	---	--	------------------------------	--	---

17:30-19:00 Posters 1 - Research track papers presented on Tuesday and Wednesday and all industrial track papers (Room: Jogging Track)

19:30-23:45 VLDB Conference Banquet at the Westin Bonaventure

Day 4: Thursday (August 29, 2019)							
Time	Room: Santa Anita	Room: San Jose	Room: San Diego	Room: Emerald Bay	Room: Hollywood Ballroom	Room: Avalon	Room: Sacramento / San Francisco
8:00-9:00	Morning Coffee (Room: California Foyer)						
9:00-10:30	VLDB 2019 Conference Awards: VLDB Endowment Awards, VLDB 10-year Best Paper Award Talk, VLDB Women in Databases Award Talk, VLDB Early Career Research Contribution Award Talk, (Room: Sacramento / San Francisco)						
10:30-11:00	Coffee Break - 30 min (Room: California Foyer)						
11:00-12:30	<b>Research Session 17:</b> Distributed Systems II	<b>Research Session 18:</b> Graph Analytics II	<b>Research Session 19:</b> Indexing and Search	<b>Research Session 20:</b> Usability and Data Mining	<b>Demo Session:</b> Group A	<b>Tutorial 5:</b> Machine Learning Meets Big Spatial Data	<b>Industry Session 3:</b> Implementing and Measuring High Performance Databases
12:30-14:00	Lunch Break - 90 min (Room: Pasadena)						
14:00-15:30	<b>Research Session 21:</b> Privacy and Security	<b>Research Session 22:</b> Clustering and Outlier Detection	<b>Research Session 23:</b> Data Integration	<b>Research Session 24:</b> Blockchain (Best Paper)	<b>Demo Session:</b> Group B	<b>Tutorial 6:</b> TextCube - Automated Construction and Multidimensional Exploration	<b>Tutorial Session 4:</b> Pretty Good Answers and Pretty Darned Fast!
15:30-16:00	Coffee Break - 30 min (Room: California Foyer)						
16:00-17:30	<b>Research Session 25:</b> Spatial Data	<b>Research Session 26:</b> Transaction Processing II	<b>Research Session 27:</b> Potpourri	<b>Research Session 28:</b> Query Processing II		<b>Tutorial 6:</b> TextCube - Automated Construction and Multidimensional Exploration	<b>Tutorial 7:</b> Personal Database Privacy and Trusted Execution Environment - A Tutorial at the Crossroads
17:30-18:30	Posters 2 - Research track papers presented on Thursday and VLDB Journal papers (Room: Jogging Track)						

Day 5: Friday (August 30, 2019)

Time	Room: Mt. Washington	Room: San Jose	Room: San Diego	Room: Santa Anita	Room: Avalon	Room: Emerald Bay & Santa Monica
9:00-10:30	CAST	DMAH	Poly	BCDL	VLIoT	BOSS
10:30-11:00	Coffee Break - 30 min (Room: California Foyer)					
11:00-12:30	CAST	DMAH	Poly	BCDL	VLIoT	BOSS
12:30-14:00	Lunch Break - 90 min (Room: Pasadena )					
14:00-15:30	CAST	DMAH	Poly	BCDL	VLIoT	BOSS
15:30-16:00	Coffee Break - 30 min (Room: California Foyer)					
16:00-17:30	CAST	DMAH	Poly	BCDL	VLIoT	BOSS



# DETAILED PROGRAM

Please note the number assigned to each research, industry, and VLDBJ paper identifies its poster stand in the Jogging track. Similarly, each number assigned to a demo paper identifies its table and easel in the Hollywood ballroom. Here are a few examples:

- For research track paper numbered 1.1, look for the stand labeled Res:1.1,
- For industrial track paper 1.1, look for the stand labeled Ind:1.1,
- For VLDB journal paper 1, look for the stand labeled VLDBJ:1,
- For Demo 5 in Group B, look for Table and Easel numbered 5.

**MONDAY – August 26 (WORKSHOPS)**

**9:00 - 18:30**

**Accelerating Analytics and Data Management Systems Using Modern Processor and Storage Architectures Workshop (ADMS)**

**Room: Emerald Bay**

*Chair: Rajesh Bordawekar (IBM T.J. Watson Research Center) and Tirthankar Lahiri (Oracle)*

**9:00 - 18:30**

**Applied AI for Database Systems and Applications (AIDB)**

**Room: San Jose**

*Chair: Berthold Reinwald (IBM Almaden Research Center), Bingsheng He (National University of Singapore) and Yingjun Wu (IBM Almaden Research Center)*

**9:00 - 18:30**

**Real-Time Business Intelligence and Analytics (BIRTE)**

**Room: San Diego**

*Chair: Malu Castellanos (Teradata Aster), Badrish Chandramouli (Microsoft Research), and Shimin Chen (Chinese Academy of Sciences)*

**9:00 - 18:30**

**PhD Workshop**

**Room: Santa Anita**

*Chair: Ilaria Bartolini (University of Bologna) and Feifei Li (Alibaba)*

**9:00 - 18:30**

**Technology Conference on Performance Evaluation and Benchmarking (TPCTC)**

**Room: Avalon**

*Chair: Raghunath Nambiar (AMD) and Meikel Poess (Oracle)*

**TUESDAY – August 27****8:00 - 8:30****Morning Coffee****Room: California Foyer****8:30 - 9:00****Conference Inauguration****Room: Sacramento / San Francisco****9:00 - 10:30***M. Tamer Özsu (University of Waterloo)***Keynote 1 : Graph Processing: A Panoramic View and Some Open Problems****Room: Sacramento / San Francisco****Chair: Lei Chen (HKUST, Hong Kong)**

**Abstract:** Graphs are not new in database community – we have long used them to model and reason about database operations and some of our earliest systems were built on graph models. There has been a renewed interest in the last decade in using graphs model real-life phenomena in many domains. Recent studies suggest that there is interest among the user community in using graphs to model applications and data that have traditionally been considered the domain of relational systems. Despite intense research and development efforts in graph processing, these efforts are fragmented, and general purpose, scalable solutions are not yet available. In this talk, I will provide a systematic look at the field and highlight some of the open research issues.



**Short Bio:** M. Tamer Özsu is a University Professor in the Cheriton School of Computer Science at the University of Waterloo where he also holds a Cheriton Faculty Fellowship. Before Waterloo, he spent 16 years at the University of Alberta. His research is on distributed data management and the management of non-conventional data. He is a Fellow of the Royal Society of Canada, American Association for the Advancement of Science, ACM and IEEE, an elected member of Academy of Science

of Turkey and a member of Sigma Xi. He is the recipient of the CS-Can/Info-Can (Canadian Computer Science Society) Lifetime Achievement Award, ACM SIGMOD Test-of-Time Award, ACM SIGMOD Contributions Award and the Ohio State University College of Engineering Distinguished Alumnus Award. He is the recipient of two best paper awards and a paper with honourable mention.

**10:30 - 11:00****Coffee Break (30 min)****Room: California Foyer****11:00 - 12:30****Research Session 1: Data-driven Machine Learning****Room: Santa Anita****Chair: Sudip Roy (Google, USA)****Res:1.1 - An Intermediate Representation for Optimizing Machine Learning Pipelines**

**Authors:** *Andreas Kunft (TU Berlin), Asterios Katsifodimos (TU Delft), Sebastian Schelter (New York University), Sebastian Breß (DFKI GmbH), Tilmann Rabl (Hasso Plattner Institute), and Volker Markl (TU Berlin)*

**Res:1.2 - An Experimental Evaluation of Large Scale GBDT Systems**

Authors: Fangcheng Fu (Peking University), Jiawei Jiang (Peking University), Yingxia Shao (BUPT), and Bin Cui (Peking University)

**Res:1.3 - Snuba: Automating Weak Supervision to Label Training Data**

Authors: Paroma Varma (Stanford University) and Christopher Re (Stanford University)

**Res:1.4 - Helix: Holistic Optimization for Accelerating Iterative Machine Learning**

Authors: Doris Xin (University of Illinois at Urbana-Champaign), Stephen Macke (University of Illinois at Urbana-Champaign), Litian Ma (University of Illinois at Urbana-Champaign), Jialin Liu (University of Illinois at Urbana-Champaign), Shuchen Song (University of Illinois at Urbana-Champaign), and Aditya Parameswaran (University of Illinois at Urbana-Champaign)

**Res:1.5 - Rafiki: Machine Learning as an Analytics Service System**

Authors: Wei Wang (National University of Singapore), Jinyang Gao (National University of Singapore), Meihui Zhang (Beijing Institute of Technology), Sheng Wang (National University of Singapore), Gang Chen (Zhejiang University), Teck Khim Ng (National University of Singapore), Beng Chin Ooi (National University of Singapore), Jie Shao (University of Electronic Science and Technology of China), and Moaz Reyad (National University of Singapore)

11:00 - 12:30

**Research Session 2: Distributed Graphs****Room: San Jose***Chair: Semih Salihoglu (University of Waterloo, Canada)***Res:2.1 - A Study of Partitioning Policies for Graph Analytics on Large-scale Distributed Platforms**

Authors: Gurbinder S Gill (The University of Texas at Austin), Roshan Dathathri (The University of Texas at Austin), Loc Hoang (The University of Texas at Austin), and Keshav Pingali (The University of Texas at Austin)

**Res:2.2 - Dynamic Scaling for Parallel Graph Computations**

Authors: Wenfei Fan (University of Edinburgh, Beihang University, Shenzhen University), Chunming Hu (Beihang Univ.), Muyang Liu (University of Edinburgh), Ping Lu (Beihang Univ.), Qiang Yin (Alibaba Group), and Jingren Zhou (Alibaba Group)

**Res:2.3 - Multi-Dimensional Balanced Graph Partitioning via Projected Gradient Descent**

Authors: Dmitrii Avdiukhin (Indiana University, Bloomington), Sergey Pupyrev (Facebook), and Grigory Yaroslavtsev (Indiana University, Bloomington)

**Res:2.4 - TopoX: Topology Refactorization for Efficient Graph Partitioning and Processing**

Authors: Dongsheng Li (National University of Defense Technology), Yiming Zhang (National University of Defense Technology), Jinyan Wang (National University of Defense Technology), and Kian-Lee Tan (National University of Singapore)

**Res:2.5 - Start Late or Finish Early: A Distributed Graph Processing System with Redundancy Reduction**

Authors: Shuang Song (The University of Texas at Austin), Xu Liu (College of William and Mary), Qinzhe Wu (The University of Texas at Austin), Andreas Gerstlauer (The University of Texas at Austin), Tao Li (University of Florida), and Lizy K. John (The University of Texas at Austin)

11:00 - 12:30

**Research Session 3: Transaction Processing I****Room: San Diego***Chair: Alan Fekete (University of Sydney, Australia)*

**Res:3.1 - Interactive Checks for Coordination Avoidance**

Authors: Michael J Whittaker (UC Berkeley) and Joseph M. Hellerstein (UC Berkeley)

**Res:3.2 - Ocean Vista: Gossip-Based Visibility Control for Speedy Geo-Distributed Transactions**

Authors: Hua Fan (Alibaba Group) and Wojciech Golab (University of Waterloo)

**Res:3.3 - Snapshot Semantics for Temporal Multiset Relations**

Authors: Anton Dignös (Free University of Bozen-Bolzano), Boris Glavic (Illinois Institute of Technology), Xing Niu (Illinois Institute of Technology), Michael H Böhlen (University of Zurich), and Johann Gamper (Free University of Bozen-Bolzano)

**Res:3.4 - SLOG: Serializable, Low-latency, Geo-replicated Transactions**

Authors: Kun Ren (eBay Inc), Dennis Li (UMD College Park), and Daniel J. Abadi (UMD College Park)

**Res:3.5 - FineLine: Log-structured Transactional Storage and Recovery**

Authors: Caetano Sauer (Tableau Software), Goetz Graefe (Google), and Theo Härder (TU Kaiserslautern)

**11:00 - 12:30****Research Session 4: Query Processing I****Room: Emerald Bay****Chair: Steffen Zeuch (DFKI Berlin, Germany)****Res:4.1 - On Obtaining Stable Rankings**

Authors: Abolfazl Asudeh (University of Michigan), H. V. Jagadish (University of Michigan), Gerome Miklau (University of Massachusetts Amherst), and Julia Stoyanovich (New York University)

**Res:4.2 - Intermittent Query Processing**

Authors: Dixin Tang (University of Chicago), Zechao Shang (University of Chicago), Aaron J Elmore (University of Chicago), Sanjay Krishnan (University of Chicago), and Michael J. Franklin (University of Chicago)

**Res:4.3 - Automated Verification of Query Equivalence Using Satisfiability Modulo Theories**

Authors: Qi Zhou (Georgia Institute of Technology), Joy Arulraj (Georgia Institute of Technology), Shamkant Navathe (Georgia Institute of Technology), William Harris (Galois Inc), and Dong Xu (Alibaba Group)

**Res:4.4 - Progressive Top-k Subarray Query Processing in Array Databases**

Authors: Dalsu Choi (Korea University), Chang-Sup Park (Dongduk Women's University), and Yon Dohn Chung (Korea University)

**Res:4.5 - Chasing similarity: distribution-aware aggregation scheduling**

Authors: Feilong Liu (The Ohio State University), Ario Salmasi (The Ohio State University), Spyros Blanas (The Ohio State University), and Anastasios Sidiropoulos (University of Illinois at Chicago)

**11:00 - 12:30****Demo Session: Group A (See page 104 for details)****Room: Hollywood Ballroom****11:00 - 12:30****Tutorial 1: Data Lake Management: Challenges and Opportunities****Room: Avalon**

**Speakers:** Fatemeh Nargesian (University of Toronto), Erkang Zhu (University of Toronto), Renée J. Miller (Northeastern University), Ken Pu (UOIT), and Patricia C. Arocena (TD Bank Group)

**Description:** The ubiquity of data lakes has created fascinating new challenges for data management research. In this tutorial, we review the state-of-the-art in data management for data lakes. We consider how data lakes are introducing new problems including dataset discovery and how they are changing the requirements for classic problems including data extraction, data cleaning, data integration, data versioning, and meta-data management. The slides of this tutorial can be downloaded at: <https://github.com/RJMillerLab/data-lake-tutorial-slides>

**11:00 - 12:30**

### Industry Session 1: Deep Learning and Graph DBs

**Room:** Sacramento / San Francisco

**Chair:** Feifei Li (Alibaba Group, China)

#### Ind:1.1 - QTune: A Query-Aware Database Tuning System with Deep Reinforcement Learning

**Authors:** Guoliang Li (Tsinghua University), Xuanhe Zhou (Tsinghua University), Shifu Li (Huawei Company), and Bo Gao (Huawei Company)

#### Ind:1.2 - Smile: A System to Support Machine Learning on EEG Data at Scale

**Authors:** Lei Cao (Massachusetts Institute of Technology), Wenbo Tao (Massachusetts Institute of Technology), Sungtae An (Georgia Institute of Technology), Jing Jin (Massachusetts General Hospital), Yizhou Yan (Massachusetts Institute of Technology), Xiaoyu Liu (Massachusetts Institute of Technology), Wendong Ge (Massachusetts General Hospital), Adam Sah (Massachusetts Institute of Technology), Leilani Battle (University of Maryland), Jimeng Sun (Georgia Institute of Technology), Remco Chang (Tufts University), Brandon Westover (Massachusetts General Hospital), Samuel Madden (Massachusetts Institute of Technology), and Michael Stonebraker (Massachusetts Institute of Technology)

#### Ind:1.3 - Guided automated learning for query workload re-optimization

**Authors:** Guilherme Damasio (Ontario Tech University, IBM Centre for Advanced Studies), Vincent Corvinelli (IBM Ltd), Parke Godfrey (York University, IBM Centre for Advanced Studies), Piotr Mierzejewski (IBM Ltd), Alex Mihaylov (Ontario Tech University, IBM Centre for Advanced Studies), Jaroslaw Szlichta (Ontario Tech University, IBM Centre for Advanced Studies), and Calisto Zuzarte (IBM Ltd)

#### Ind:1.4 - AliGraph: A Comprehensive Graph Neural Network Platform

**Authors:** Rong Zhu (Alibaba Group), Kun Zhao (Alibaba Group), Hongxia Yang (Alibaba Group), Wei Lin (Alibaba Group), Chang Zhou (Alibaba Group), Baole Ai (Alibaba Group), Yong Li (Alibaba Group), and Jingren Zhou (Alibaba Group)

#### Ind:1.5 - Updating Graph Databases with Cypher

**Authors:** Alastair Green (Neo4j), Paolo Guagliardo (University of Edinburgh), Leonid Libkin (University of Edinburgh), Tobias Lindaaker (Neo4j), Victor Marsault (LIGM, UPEM/ESIEE-Paris/ENPC/CNRS), Stefan Plantikow (Neo4j), Martin Schuster (Abbott Informatics), Petra Selmer (Neo4j), and Hannes Voigt (Neo4j)

**12:30 - 14:00**

### Lunch Break (90 min)

**Room:** Pasadena

*Tova Milo (Tel Aviv University, Israel)*

**14:00 - 15:30**

### Keynote 2 : Getting Rid of Data

**Room:** Sacramento / San Francisco

**Chair: Divesh Srivastava (AT&T Labs-Research, USA)**

**Abstract:** We are experiencing an amazing data-centered revolution. Incredible amounts of data are collected, integrated and analyzed, leading to key breakthroughs in science and society. This well of knowledge, however, is at a great risk if we do not dispense with some of the data flood. First, the amount of generated data grows exponentially and already at 2020 is expected to be more than twice the available storage. Second, even disregarding storage constraints, uncontrolled data retention risks privacy and security, as recognized, e.g., by the recent EU Data Protection reform. Data disposal policies must be developed to benefit and protect organizations and individuals. Retaining the knowledge hidden in the data while respecting storage, processing and regulatory constraints is a great challenge. The difficulty stems from the distinct, intricate requirements entailed by each type of constraint, the scale and velocity of data and the constantly evolving needs. While multiple data sketching, summarization and deletion techniques were developed to address specific aspects of the problem, we are still very far from a comprehensive solution. Every organization has to battle the same tough challenges, with ad hoc solutions that are application specific and rarely sharable. In this talk I will discuss the logical, algorithmic, and methodological foundations required for the systematic disposal of large-scale data, for constraints enforcement and for the development of applications over the retained information. I will overview relevant related work, highlighting new research challenges and potential reuse of existing techniques, as well as the research performed in this direction in the Tel Aviv Databases group.



**Short Bio:** Tova Milo received her Ph.D. degree in Computer Science from the Hebrew University, Jerusalem, in 1992. After graduating she worked at the INRIA research institute in Paris and at University of Toronto and returned to Israel in 1995, joining the School of Computer Science at Tel Aviv university, where she is now a full Professor and holds the Chair of Information Management. She served as the Head of the Computer Science Department from 2011-2014. Her research focuses on

large-scale data management applications such as data integration, semi-structured information, Data-centered Business Processes and Crowd-sourcing, studying both theoretical and practical aspects. Tova served as the Program Chair of multiple international conferences, including PODS, VLDB, ICDT, XSym, and WebDB, and as the chair of the PODS Executive Committee. She served as a member of the VLDB Endowment and the PODS and ICDT executive boards and as an editor of TODS, IEEE Data Eng. Bull, and the Logical Methods in Computer Science Journal. Tova has received grants from the Israel Science Foundation, the US-Israel Binational Science Foundation, the Israeli and French Ministry of Science and the European Union. She is an ACM Fellow, a member of Academia Europaea, a recipient of the 2010 ACM PODS Alberto O. Mendelzon Test-of-Time Award, the 2017 VLDB Women in Database Research award, the 2017 Weizmann award for Exact Sciences Research, and of the prestigious EU ERC Advanced Investigators grant.

**15:30 - 16:00****Coffee Break (30 min)****Room: California Foyer****16:00 - 17:30****Research Session 5: Cloud Databases****Room: Santa Anita****Chair: Umar Farooq Minhas (Microsoft Research, USA)****Res:5.1 - Autoscaling Tiered Cloud Storage in Anna**

**Authors:** Chenggang Wu (UC Berkeley), Vikram Sreekanti (UC Berkeley), and Joseph M Hellerstein (UC Berkeley)

**Res:5.2 - iBTune: Individualized Buffer Tuning for Large-scale Cloud Databases**

*Authors: Jian Tan (Alibaba), Teying Zhang (Alibaba Group), Feifei Li (Alibaba), Jie Chen (Alibaba), Qixing Zheng (Alibaba), Ping Zhang (Alibaba), Honglin Qiao (Alibaba), Yue Shi (Alibaba), Wei Cao (Alibaba Cloud), and Rui Zhang (Alibaba)*

**Res:5.3 - Unifying Consensus and Atomic Commitment for Effective Cloud Data Management**

*Authors: Sujaya A Maiyya (University Of California, Santa Barbara), Faisal Nawab (UC Santa Cruz), Divy Agrawal (University of California, Santa Barbara), and Amr El Abbadi (UC Santa Barbara)*

**Res:5.4 - Pangea: Monolithic Distributed Storage for Data Analytics**

*Authors: Jia Zou (Rice University), Arun Iyengar (IBM T.J. Watson Research Center), and Chris Jermaine (Rice University)*

**Res:5.5 - An IDEA: An Ingestion Framework for Data Enrichment in AsterixDB**

*Authors: Xikui Wang (University of California, Irvine) and Michael Carey (University of California, Irvine)*

16:00 - 17:30

**Research Session 6: Subgraphs and Communities****Room: San Jose***Chair: A. Erdem Sariyuce (University at Buffalo, USA)***Res:6.1 - Distributed Subgraph Matching on Timely Dataflow**

*Authors: Longbin Lai (UNSW), Zhu Qing (East China Normal University), Zhengyi Yang (University of New South Wales, Sydney), Xin Jin (East China Normal University), Zhengmin Lai (East China Normal University), Ran Wang (East China Normal University), Kongzhang Hao (UNSW), Xuemin Lin (University of New South Wales), Lu Qin (UTS), Wenjie Zhang (University of New South Wales), Ying Zhang (University of Technology Sydney), Zhengping Qian (Alibaba Group), and Jingren Zhou (Alibaba Group)*

**Res:6.2 - Local Algorithms for Hierarchical Dense Subgraph Discovery**

*Authors: A. Erdem Sariyuce (University at Buffalo, US), C. Seshadhri (UCSC), and Ali Pinar (Sandia National Laboratories)*

**Res:6.3 - Fast and Robust Distributed Subgraph Enumeration**

*Authors: Xuguang Ren (Griffith University), Junhu Wang (Griffith University), Wook-Shin Han (POSTECH), and Jeffrey Xu Yu (The Chinese University of Hong Kong)*

**Res:6.4 - Finding Theme Communities from Database Networks**

*Authors: Lingyang Chu (Simon Fraser University), Zhefeng Wang (Huawei Technologies), Jian Pei (Simon Fraser University), Yanyan Zhang (Simon Fraser University), Yu Yang (Simon Fraser University), and Enhong Chen (University of Science and Technology of China)*

**Res:6.5 - Vertex Priority Based Butterfly Counting for Large-scale Bipartite Networks**

*Authors: Kai Wang (University of New South Wales), Xuemin Lin (University of New South Wales), Lu Qin (University of Technology Sydney), Wenjie Zhang (University of New South Wales), and Ying Zhang (University of Technology Sydney)*

16:00 - 17:30

**Research Session 7: Similarity Search and Data Extraction****Room: San Diego***Chair: Chen Li (University of California Irvine, USA)***Res:7.1 - Towards a Unified Framework for String Similarity Joins**

*Authors: Pengfei Xu (University of Helsinki) and Jiaheng Lu (University of Helsinki)*



**Res:7.2 - Pigeonring: A Principle for Faster Thresholded Similarity Search**

*Authors: Jianbin Qin (The University of Edinburgh) and Chuan Xiao (Nagoya University)*

**Res:7.3 - The Lernaean Hydra of Data Series Similarity Search: An Experimental Evaluation of the State of the Art**

*Authors: Karima Echihabi (ENSIAS, Université Mohammed V de Rabat), Kostas Zoumpatianos (Harvard University), Themis Palpanas (Paris Descartes University), and Houda Benbrahim (ENSIAS, Université Mohammed V de Rabat)*

**Res:7.4 - Balance-Aware Distributed String Similarity-Based Query Processing System**

*Authors: Ji Sun (Tsinghua University), Zeyuan Shang (MIT), Guoliang Li (Tsinghua University), Dong Deng (MIT), and Zhifeng Bao (RMIT University)*

**Res:7.5 - Online Template Induction for Machine-Generated Emails**

*Authors: Michael J Whittaker (UC Berkeley), Nick Edmonds (Google), Sandeep Tata (Google), James B Wendt (Google), and Marc Najork (Google)*

**16:00 - 17:30****Research Session 8: Hardware and Machine Learning****Room: Emerald Bay**

*Chair: Berthold Reinwald (IBM Research - Almaden, USA)*

**Res:8.1 - HetExchange: Encapsulating heterogeneous CPU-GPU parallelism in JIT compiled engines**

*Authors: Periklis Chrysogelos (EPFL), Manos Karpathiotakis (Facebook), Raja Appuswamy (Eurecom), and Anastasia Ailamaki (EPFL)*

**Res:8.2 - Analyzing Efficient Stream Processing on Modern Hardware**

*Authors: Steffen Zeuch (DFKI GmbH), Bonaventura Del Monte (DFKI GmbH), Jeyhun Kari-mov (DFKI GmbH), Clemens Lutz (DFKI GmbH), Manuel Renz (DFKI GmbH), Jonas Traub (Technische Universität Berlin), Sebastian Breß (DFKI GmbH), Tilmann Rabl (DFKI GmbH), and Volker Markl (Technische Universität Berlin)*

**Res:8.3 - Accelerating Generalized Linear Models with MLWeaving: A One-Size-Fits-All System for Any-precision Learning**

*Authors: Zeke Wang (ETH Zurich), Kaan Kara (ETH Zurich), Hantian Zhang (ETH Zurich), Gustavo Alonso (ETH Zurich), Onur Mutlu (ETH Zurich), and Ce Zhang (ETH Zurich)*

**Res:8.4 - Crossbow: Scaling Deep Learning with Small Batch Sizes on Multi-GPU Servers**

*Authors: Alexandros Koliouisis (Imperial College London), Pijika Watcharapichat (Microsoft Research), Matthias Weidlich (Humboldt-Universität zu Berlin), Luo Mai (Imperial College London), Paolo Costa (Microsoft Research), and Peter Pietzuch (Imperial College London)*

**Res:8.5 - ColumnML: Column-Store Machine Learning with On-The-Fly Data Transformation**

*Authors: Kaan Kara (ETH Zurich), Ken Eguro (Microsoft), Ce Zhang (ETH Zurich), and Gustavo Alonso (ETH Zurich)*

**16:00 - 17:30****Demo Session: Group B (See page 104 for details)****Room: Hollywood Ballroom****16:00 - 17:30****Tutorial 2: Combating Fake News: A Data Management and Mining Perspective****Room: Avalon**

*Speakers: Laks V.S. Lakshmanan (The University of British Columbia), Michael Simpson (The University of British Columbia), and Saravanan Thirumuruganathan (QCRI, HBKU)*

**Description:** Fake news is a major threat to global democracy resulting in diminished trust in government, journalism and civil society. The public popularity of social media and social networks has caused a contagion of fake news where conspiracy theories, disinformation and extreme views flourish. Detection and mitigation of fake news is one of the fundamental problems of our times and has attracted widespread attention. While fact checking websites such as snopes, politifact and major companies such as Google, Facebook, and Twitter have taken preliminary steps towards addressing fake news, much more remains to be done. As an interdisciplinary topic, various facets of fake news have been studied by communities as diverse as machine learning, databases, journalism, political science and many more. The objective of this tutorial is two-fold. First, we wish to familiarize the database community with the efforts by other communities on combating fake news. We provide a panoramic view of the state-of-the-art of research on various aspects including detection, propagation, mitigation, and intervention of fake news. Next, we provide a concise and intuitive summary of prior research by the database community and discuss how it could be used to counteract fake news. The tutorial covers research from areas such as data integration, truth discovery and fusion, probabilistic databases, knowledge graphs and crowdsourcing from the lens of fake news. Effective tools for addressing fake news could only be built by leveraging the synergistic relationship between database and other research communities. We hope that our tutorial provides an impetus towards such synthesis of ideas and the creation of new ones. The slides for this tutorial can be downloaded at <https://combatingfakenewstutorial.github.io/v1db19.html>

**16:00 - 17:30**

**Invited Industry Talks 1: Data Standing Tall in the Clouds**

**Room: Sacramento / San Francisco**

**Chair: Beng Chin Ooi (NUS, Singapore)**

*FeiFei Li (Alibaba)*

**Cloud Native Database Systems at Alibaba: Opportunities and Challenges**

**Abstract:** Cloud native databases become increasingly important for the era of cloud computing, due to the needs for elasticity and on-demand usage by various applications. These challenges from cloud applications present new opportunities for cloud native databases that cannot be fully addressed by traditional on-premise enterprise database systems. A cloud native database leverages software-hardware co-design to explore accelerations offered by new hardware such as RDMA, NVM, kernel bypassing protocols such as SDPK and DPDK, etc. Meanwhile, new design architectures such as shared-storage and shared-everything enable a cloud-native database to decouple computation from storage and provide excellent elasticity. For highly concurrent workloads that require horizontal scalability, a cloud native database can also leverage a shared-nothing layer to provide distributed query and transaction processing. Applications also require that cloud native databases to offer high availability through distributed consensus protocols. At Alibaba, we have explored a suite of technologies to design cloud-native database systems. Our storage engine, X-engine and PolarFS, improves both write and read throughputs by using a LSM-tree design and self-adapted separation of hot and cold data records. Based on these efforts, we have designed and implemented PolarDB and its distributed version PolarDB-X, which has successfully supported the extreme transaction workloads during the 2018 Global Shopping Festival on November 11, 2018, and achieved commercial success on Alibaba cloud. In addition, we have explored a self-driving database platform to achieve autoscaling and intelligent database management. We will report key technologies and lessons learned to highlight the technical challenges and opportunities for cloud native database systems at Alibaba.



**Short Bio:** Feifei Li is currently a Vice President of Alibaba Group, ACM Distinguished Scientist, Director of the Database Products Business Unit of Alibaba Cloud and the Database and Storage Lab of DAMO academy. He is also a full professor at the School of Computing, University of Utah (on leave). He has won multiple awards from NSF, ACM, IEEE, Visa, Google, and others. In particular, he is a recipient of the IEEE ICDE 2014 10 Years Most Influential Paper Award, ACM SIGMOD 2016 Best Paper Award, ACM SIGMOD 2015 Best System Demonstration Award, IEEE ICDE 2004 Best Paper Award, NSF Career Award, Google Faculty

award, and Visa Faculty award. He has been an associate editor, co-chairs, and core committee members for many prestigious research journals and conferences.

*Dharma Shukla (Microsoft)*

### **Azure Cosmos DB: Microsoft's Globally Distributed Database Service**

**Abstract:** As a cloud service, Cosmos DB is carefully architected and engineered with multi-tenancy and global distribution from the ground up. Cosmos DB allows customers to add/remove any number of Azure regions to their Cosmos database, at any time. The service allows customers to elastically scale the throughput and storage of their Cosmos database, in an unlimited manner. By virtue of its novel multi-master replication protocol, the service offers guaranteed single-digit millisecond low latencies for writes and reads at P99 along with 99.999 high-availability. Cosmos DB allows developers to choose from five well-defined consistency models. Cosmos DB is used extensively both internally within Microsoft as well as externally by developers using Microsoft Azure. In this talk, Dharma will describe the design of Cosmos DB as well as his learnings from designing and operating a globally distributed service across heterogeneous customer workloads worldwide. Along with several key design ideas, Dharma will also present several crucial engineering choices they had to make.



**Short Bio:** Dharma Shukla is a Technical Fellow at Microsoft. Dharma is also the founder and architect of Azure Cosmos DB. You can reach out to Dharma at [dharmas@microsoft.com](mailto:dharmas@microsoft.com) or [@dharmashukla](https://twitter.com/dharmashukla) on Twitter.

*Mike Carey (UCI and Couchbase)*

### **Couchbase Analytics: NoETL for Scalable NoSQL Data Analysis**

**Abstract:** Couchbase Server is a highly scalable document-oriented database management system. With a shared-nothing architecture, it exposes a fast key-value store with a managed cache for sub-millisecond data operations, indexing for fast queries, and a powerful query engine for executing declarative SQL-like queries. Its Query Service debuted several years ago and supports high volumes of low-latency queries and updates for JSON documents. Its recently introduced Analytics Service complements the Query Service. Couchbase Analytics, the focus of this talk, supports complex analytical queries (e.g., ad hoc joins and aggregations) over large collections of JSON documents. This talk will describe the Analytics service from the outside in, including its user model, its SQL++ based query language, and its MPP-based storage and query processing architecture. It will also touch on the relationship of Couchbase Analytics to Apache AsterixDB, the open source Big Data management system at the core of Couchbase Analytics.



**Short Bio:** Michael Carey received his B.S. and M.S. degrees from Carnegie-Mellon University and his Ph.D. from the University of California, Berkeley. He is currently a Bren Professor of Information and Computer Sciences and Distinguished Professor of Computer Science at UC Irvine, where he leads the AsterixDB project, as well as a Consulting Architect at Couchbase, Inc. Before joining UCI in 2008, he worked at BEA Systems for seven years and led the development of their Aqua-Logic Data Services Platform product for virtual data integration. He also spent a dozen years at the University of Wisconsin-Madison, five years at the IBM Almaden Research Center working on object-relational databases, and a year and a half at e-commerce platform startup Propel Software during the infamous 2000-2001 Internet bubble. He is an ACM Fellow, an IEEE Fellow, a member of the National Academy of Engineering, and a recipient of the ACM SIGMOD E.F. Codd Innovations Award. His current interests center around data-intensive computing and scalable data management (a.k.a. Big Data).

**17:45-18:30**

**Departure for the Reception, California Science Center**

**19:00 - 22:00**

**VLDB Conference Reception at the California Science Center**

## **WEDNESDAY – August 28**

**8:00 - 8:30**

**Morning Coffee**

**Room: California Foyer**

**8:30 - 9:00**

**DBCares (Chaudhuri), VLDB Endowment (Markl), VLDB 2020 (Ishikawa)**

**Room: Sacramento / San Francisco**

*Eric Iverson (CIO, Creative Artists Agency)*

**Keynote 3 : Awesome Data Use Cases from Hollywood:  
How data is changing everything about the entertainment  
experience**

**9:00 - 10:30**

**Room: Sacramento / San Francisco**

**Chair: Fatma Özcan (IBM Research-Almaden, USA)**

**Abstract:** For more than 75 years, movies and television have engaged our imagination with gripping stories highlighting the impact of technology on the world and the worlds we create. And while Hollywood has always found the importance of data and technology in storytelling, its role in the business has been persistent, but understated. That has all changed. From music to sports to TV to the movies, data is changing everything about how entertainment works. Data now plays important roles in what content we make, where we show it, how we make it, who sees it, and how we can improve the overall experience when audiences engage with entertainment. In this session we will explore the surprising and fascinating ways data is changing all aspects Hollywood and the entertainment industry both now and into the future.



**Short Bio:** Eric Iverson is the Chief Information Officer (CIO) at leading entertainment and sports agency Creative Artists Agency (CAA). Iverson is based in the Los Angeles office and oversees the Information Technology department. His responsibilities include IT technology support, cybersecurity, data analytics engineering, application solutions, and digital media solutions. Prior to joining CAA in 2016, Iverson spent more than 17 years as CIO at Sony Pictures Television where he was responsible for the global systems portfolio and led strategic technology change

initiatives for television production, distribution, and network channels. He graduated from Pepperdine University with a degree in music composition.

**10:30 - 11:00**

**Coffee Break (30 min)**

**Room: California Foyer**

**11:00 - 12:30**

**Research Session 9: Distributed Systems I**

**Room: Santa Anita**

**Chair: Hakan Hacigumus (Google, USA)**

**Res:9.1 - Hillview: A trillion-cell spreadsheet for big data**

**Authors:** Mihai Budiu (VMware Research), Parikshit Gopalan (VMware Research), Lalith Suresh (VMware Research), Udi Wieder (VMware Research), Han Kruijer (University of Groningen), and Marcos K. Aguilera (VMware Research)

**Res:9.2 - Yugong: Geo-Distributed Data and Job Placement at Scale**

**Authors:** Yuzhen Huang (The Chinese University of Hong Kong, Alibaba Group), Yingjie Shi (Alibaba Group), Zheng Zhong (The Chinese University of Hong Kong, Alibaba Group), Yihui Feng (Alibaba Group), James Cheng (The Chinese University of Hong Kong), Jiwei Li (Alibaba Group), Haochuan Fan (Alibaba Group), Chao Li (Alibaba Group), Tao Guan (Alibaba Group), and Jingren Zhou (Alibaba Group)

**Res:9.3 - Hyper Dimension Shuffle: Efficient Data Repartition at Petabyte Scale in SCOPE**

**Authors:** Shi Qiao (Microsoft), Adrian Nicoara (Microsoft), Jin Sun (Microsoft), Marc T. Friedman (Microsoft), Hiren Patel (Microsoft), and Jaliya Ekanayake (Microsoft)

**Res:9.4 - Declarative Recursive Computation on an RDBMS, or, Why You Should Use a**

**Database For Distributed Machine Learning**

*Authors: Dimitrije Jankov (Rice University), Shangyu Luo (Rice University), Binhang Yuan (Rice University), Zhuhua Cai (Rice University), Jia Zou (Rice University), Chris Jermaine (Rice University), and Zekai J Gao (Rice University)*

**Res:9.5 - Procella: Unifying serving and analytical data at YouTube**

*Authors: Biswapesh Chattopadhyay (Google LLC), Priyam Dutta (Google LLC), Weiran Liu (Google LLC), Ott Tinn (Google LLC), Andrew McCormick (Google LLC), Aniket Mokashi (Google LLC), Paul Harvey (Google LLC), Hector Gonzalez (Google LLC), David Lomax (Google LLC), Sagar Mittal (Google LLC), Roei Ebenstein (Google LLC), Nikita Mikhaylin (Google LLC), Hung-ching Lee (Google LLC), Xiaoyan Zhao (Google LLC), Tony Xu (Google LLC), Luis Perez (Google LLC), Farhad Shahmohammadi (Google LLC), Tran Bui (Google LLC), Neil McKay (Google LLC), Selcuk Aya (Google LLC), Vera Lychagina (Google LLC), and Brett Elliott (Google LLC)*

**11:00 - 12:30****Research Session 10: Graph Analytics I****Room: San Jose****Chair: Arijit Khan (Nanyang Technological University, Singapore)****Res:10.1 - Deducing Certain Fixes to Graphs**

*Authors: Wenfei Fan (Univ. of Edinburgh & Beihang Univ.), Ping Lu (Beihang Univ.), Chao Tian (Alibaba Group), and Jingren Zhou (Alibaba Group)*

**Res:10.2 - Certus: An Effective Entity Resolution Approach with Graph Differential Dependencies (GDDs)**

*Authors: Selasi Kwashie (University of South Australia), Jixue Liu (University of South Australia), Jiuyong Li (University of South Australia), Lin Liu (University of South Australia), Markus Stumptner (University of South Australia), and Lujing Yang (University of South Australia)*

**Res:10.3 - Efficient Algorithms for Densest Subgraph Discovery**

*Authors: Yixiang Fang (Guangzhou University), Kaiqiang Yu (The University of Hong Kong), Reynold Cheng (The University of Hong Kong), Laks V.S. Lakshmanan (The University of British Columbia), and Xuemin Lin (University of New South Wales)*

**Res:10.4 - Beyond Macrobenchmarks: Microbenchmark-based Graph Database Evaluation**

*Authors: Matteo Lissandrini (Aalborg University), Martin Brugnara (University of Trento), and Yannis Velegrakis (Utrecht University)*

**Res:10.5 - Utility-Driven Graph Summarization**

*Authors: K. Ashwin Kumar (Symantec Research Labs) and Petros Efsthathopoulos (Symantec Research Labs)*

**11:00 - 12:30****Research Session 11: Stream Processing and Analysis****Room: San Diego****Chair: Nick Koudas (University of Toronto, Canada)****Res:11.1 - Stream Frequency Over Interval Queries**

*Authors: Ran Ben Basat (Harvard University), Roy Friedman (Technion), and Rana Shahout (Technion)*

**Res:11.2 - SWIFT: Mining Representative Patterns from Large Event Streams**

*Authors: Yizhou Yan (Worcester Polytechnic Institute), Lei Cao (MIT), Samuel Madden (MIT), and Elke Rundensteiner (Worcester Polytechnic Institute)*

**Res:11.3 - Optimal and General Out-of-Order Sliding-Window Aggregation**

*Authors: Kanat Tangwongsan (Mahidol University International College), Martin Hirzel (IBM Research), and Scott Schneider (IBM Research)*

**Res:11.4 - From Anomaly Detection to Rumour Detection using Data Streams of Social Platforms**

*Authors: Thanh Tam Nguyen (Ecole Polytechnique Federale de Lausanne), Matthias Weidlich (Humboldt-Universität zu Berlin), Bolong Zheng (Huazhong University of Science and Technology), Hongzhi Yin (The University of Queensland), Nguyen Quoc Viet Hung (Griffith University), and Bela Stantic (Griffith University)*

**Res:11.5 - Real-time Distributed Co-Movement Pattern Detection on Streaming Trajectories**

*Authors: Lu Chen (Aalborg University, Denmark), Yunjun Gao (Zhejiang University), Ziquan Fang (Zhejiang University), Xiaoye Miao (Zhejiang University), Christian S Jensen (Aalborg University), and Chenjuan Guo (Aalborg University)*

**11:00 - 12:30****Research Session 12: Database Engines for New Hardware****Room: Emerald Bay**

*Chair: Spyros Blanas (The Ohio State University, USA)*

**Res:12.1 - Generating Application-specific Data Layouts for In-memory Databases**

*Authors: Cong Yan (University of Washington) and Alvin Cheung (UC Berkeley)*

**Res:12.2 - DimmStore: Memory Power Optimization for Database Systems**

*Authors: Alexey Karyakin (University of Waterloo) and Kenneth Salem (University of Waterloo)*

**Res:12.3 - Accelerating Raw Data Analysis with the ACCORDA Software and Hardware Architecture**

*Authors: Yuanwei Fang (University of Chicago), Chen Zou (University of Chicago), and Andrew Chien (University of Chicago)*

**Res:12.4 - Everything You Always Wanted to Know About Compiled and Vectorized Queries But Were Afraid to Ask**

*Authors: Timo Kersten (Technische Universität München), Viktor Leis (Technische Universität München), Alfons Kemper (Technische Universität München), Thomas Neumann (Technische Universität München), Andrew Pavlo (Carnegie Mellon University), and Peter Boncz (Centrum Wiskunde & Informatica)*

**Res:12.5 - Rethinking Database High Availability with RDMA Networks**

*Authors: Erfan Zamanian (Brown University), Xiangyao Yu (MIT), Michael Stonebraker (MIT), and Tim Kraska (MIT)*

**11:00 - 12:30****Demo Session: Group A (See page 104 for details)****Room: Hollywood Ballroom****11:00 - 12:30****Tutorial 3: Speedup Your Analytics: Automatic Parameter Tuning for Databases and Big Data Systems****Room: Avalon**

*Speakers: Jiaheng Lu (University of Helsinki), Yuxing Chen (University of Helsinki), Herodotos Herodotou (Cyprus University of Technology), and Shivnath Babu (Duke University)*

*Description:* Database and big data analytics systems such as Hadoop and Spark have a large number of configuration parameters that control memory distribution, I/O optimization, paral-

lelism, and compression. Improper parameter settings can cause significant performance degradation and stability issues. However, regular users and even expert administrators struggle to understand and tune them to achieve good performance. In this tutorial, we review existing approaches on automatic parameter tuning for databases, Hadoop, and Spark, which we classify into six categories: rule-based, cost modeling, simulation-based, experiment-driven, machine learning, and adaptive tuning. We describe the foundations of different automatic parameter tuning algorithms and present pros and cons of each approach. The slides of this tutorial can be downloaded at: <https://www.helsinki.fi/en/researchgroups/unified-database-management-systems-udbms/tutorial/vldb-2019-tutorial>

**11:00 - 12:30**

**Industry Session 2: High Performance DBs: Both General Purpose and Specialized**

**Room: Sacramento / San Francisco**

**Chair: Alex Boehm (SAP SE)**

**Ind:2.1 - A Lightweight and Efficient Temporal Database Management System in TDSQL**

*Authors: Wei Lu (Renmin University of China), Zhanhao Zhao (Renmin University of China), Xiaoyu Wang (Tencent Inc.), Haixiang Li (Tencent Inc.), Zhenmiao Zhang (Renmin University of China), Zhiyu Shui (Renmin University of China), Sheng Ye (Tencent Inc.), Anqun Pan (Tencent Inc.), and Xiaoyong Du (Renmin University of China)*

**Ind:2.2 - TitAnt: Online Real-time Transaction Fraud Detection in Ant Financial**

*Authors: Shaosheng Cao (Ant Financial Services Group), XinXing Yang (Ant Financial Services Group), Cen Chen (Ant Financial Services Group), Jun Zhou (Ant Financial Services Group), Xiaolong Li (Ant Financial Services Group), and Yuan Qi (Ant Financial Services Group)*

**Ind:2.3 - A Distributed System for Large-scale n-gram Language Models at Tencent**

*Authors: Qiang Long (Tencent), Wei Wang (National University of Singapore), Jinfu Deng (Tencent), Song Liu (Tencent), Wenhao Huang (Tencent), Fangying Chen (Tencent), and Sifan Liu (Tencent)*

**Ind:2.4 - AnalyticDB: Real-time OLAP Database System at Alibaba Cloud**

*Authors: Chaqun Zhan (Alibaba Group), Maomeng Su (Alibaba Group), Chuangxian Wei (Alibaba Group), Xiaoqiang Peng (Alibaba Group), Liang Lin (Alibaba Group), Sheng Wang (Alibaba Group), Zhe Chen (Alibaba Group), Feifei Li (Alibaba Group), Yue Pan (Alibaba Group), Fang Zheng (Alibaba Group), and Chengliang Chai (Alibaba Group)*

**Ind:2.5 - Constant Time Recovery in Azure SQL Database**

*Authors: Panagiotis Antonopoulos (Microsoft), Peter Byrne (Microsoft), Wayne Chen (Microsoft), Cristian Diaconu (Microsoft), Raghavendra Thallam Kodandaramaiah (Microsoft), Hanuma Kodavalla (Microsoft), Prashanth Purnananda (Microsoft), Adrian-Leonard Radu (Microsoft), Chaitanya Sreenivas Ravella (Microsoft), and Girish Mittur Venkataramanappa (Microsoft)*

**12:30 - 14:00**

**Lunch on your own (90 min)**

**14:00 - 15:30**

**Panel: Opportunities for data-management research in the era of horizontal AI/ML**

**Room: Sacramento / San Francisco**

**Moderator: Theo Rekatsinas (University of Wisconsin)**

**Participants:** Theo Rekatsinas (University of Wisconsin), Ce Zhang (ETH Zurich), Sudeepa Roy (Duke University), Manasi Vertak (Verta.AI), Neoklis (Alkis) Polyzotis (Google Research)



**Abstract:** AI/ML is becoming a horizontal technology: its application is expanding to more domains, and its integration touches more parts of the technology stack. Given the strong dependence of ML on data, this expansion creates a new space for applying data management techniques. At the same time, the deeper integration of ML in the technology stack provides more touch points where ML can be used in data management systems and vice versa. In this panel, we invite researchers working in this domain to discuss this emerging world and its implications on data-management research. Among other topics, the discussion will touch on the opportunities for interesting research, how we can interact with other communities, what is the core expertise we bring to the table, and how we can conduct and evaluate this research effectively within our own community. The goal of the panel is to nudge the community to appreciate the opportunities in this new world of horizontal AI/ML and to spur a discussion on how we can shape an effective research agenda.

**15:30 - 16:00**

**Coffee Break (30 min)**

**Room: California Foyer**

**16:00 - 17:30**

**Research Session 13: Query Optimization**

**Room: Santa Anita**

**Chair: Johann-Christoph Freytag (Humboldt-Universität zu Berlin, Germany)**

**Res:13.1 - Towards a Learning Optimizer for Shared Clouds**

*Authors: Chenggang Wu (UC Berkeley), Alekh Jindal (Microsoft), Saeed Amizadeh (Microsoft), Hiren Patel (Microsoft), Wangchao le (Microsoft), Shi Qiao (Microsoft), and Sriram Rao (Facebook)*

**Res:13.2 - Plan-Structured Deep Neural Network Models for Query Performance Prediction**

*Authors: Ryan C Marcus (Brandeis University) and Olga Papaemmanouil (Brandeis University)*

**Res:13.3 - Neo: A Learned Query Optimizer**

*Authors: Ryan C Marcus (Brandeis University), Parimarjan Negi (MIT), Hongzi Mao (MIT), Chi Zhang (Brandeis University), Mohammad Alizadeh (MIT CSAIL), Tim Kraska (MIT), Olga Papaemmanouil (Brandeis University), and Nesime Tatbul (Intel Labs and MIT)*

**Res:13.4 - Comparing Synopsis Techniques for Approximate Spatial Data Analysis**

*Authors: A. B. Siddique (University of California, Riverside), Ahmed Eldawy (University of California, Riverside), and Vagelis Hristidis (UC Riverside)*

**16:00 - 17:30**

**Research Session 14: Data Integration and Crowdsourcing**

**Room: San Jose**

**Chair: Letzia Tanca (Politecnico di Milano, Italy)**

**Res:14.1 - Meta-Mappings for Schema Mapping Reuse**

*Authors: Paolo Atzeni (Roma Tre University), Luigi Bellomarini (Banca d'Italia), Paolo Papotti (Eurecom), and Riccardo Torlone (Roma Tre University)*

**Res:14.2 - Cost-efficient Data Acquisition on Online Data Marketplaces for Correlation Analysis**

*Authors: Yanying Li (Stevens Institute of Technology), Haipei Sun (Stevens Institute of Technology), Boxiang Dong (Montclair State University), and Hui (Wendy) Wang (Stevens Institute of Technology)*

**Res:14.3 - Smurf: Self-Service String Matching Using Random Forests**

*Authors: Paul Suganthan G C (University of Wisconsin-Madison), Adel Ardalan (University of Wisconsin-Madison), AnHai Doan (University of Wisconsin-Madison), and Aditya Akella (University of Wisconsin-Madison)*

**Res:14.4 - Cost-Effective Data Annotation using Game-Based Crowdsourcing**

*Authors: Jingru Yang (Renmin University of China), Ju Fan (Renmin University of China), Zhewei Wei (Renmin University of China), Guoliang Li (Tsinghua University), Tongyu Liu (Renmin University of China), and Xiaoyong Du (Renmin University of China)*

**Res:14.5 - Cleaning Crowdsourced Labels Using Oracles For Statistical Classification**

*Authors: Mohamad Dolatshah (Simon Fraser University), Mathew Teoh (Simon Fraser University), Jiannan Wang (Simon Fraser University), and Jian Pei (JD.com, Simon Fraser University)*

**16:00 - 17:30****Research Session 15: New Architectures and Caching****Room: San Diego****Chair: Mohamed Eltabakh (WPI, USA)****Res:15.1 - Efficient Data Ingestion and Query Processing for LSM-Based Storage Systems**

*Authors: Chen Luo (University of California, Irvine) and Michael Carey (University of California, Irvine)*

**Res:15.2 - DASH: Database Shadowing for Mobile DBMS**

*Authors: Youjip Won (KAIST), Sundoo Kim (Hanyang University), Juseong Yun (Hanyang University), DamQuang Tuan (Hanyang University), and Jiwon Seo (Hanyang University)*

**Res:15.3 - Block as a Value for SQL over NoSQL**

*Authors: Yang Cao (University of Edinburgh), Wenfei Fan (Univ. of Edinburgh & Beihang Univ.), and Tengfei Yuan (University of Edinburgh)*

**Res:15.4 - Cache-aware load balancing of data center applications**

*Authors: Aaron Archer (Google), Kevin Aydin (Google), MohammadHossein Bateni (Google), Vahab Mirrokni (Google), Aaron Schild (UC Berkeley), Ray Yang (Google), and Richard Zhuang (Google)*

**Res:15.5 - Design, Implementation, and Evaluation of Write-Back Policy with Cache Augmented Data Stores**

*Authors: Shahram Ghandeharizadeh (University of Southern California) and Hieu T Nguyen (University of Southern California)*

**16:00 - 17:30****Research Session 16: Data Exploration****Room: Emerald Bay****Chair: Fabio Porto (LNCC, USA)****Res:16.1 - Optimization for Active Learning-based Interactive Database Exploration**

*Authors: Enhui Huang (Ecole Polytechnique), Liping Peng (University of Massachusetts Amherst), Luciano Di Palma (Ecole Polytechnique), Ahmed Abdelkafi (Ecole Polytechnique), Anna Liu (University of Massachusetts Amherst), and Yanlei Diao (Ecole Polytechnique and University of Massachusetts Amherst)*

**Res:16.2 - DIFF: A Relational Interface for Large-Scale Data Explanation**

*Authors: Firas Abuzaaid (Stanford University), Peter Kraft (Stanford University), Sahaana Suri (Stanford University), Edward Gan (Stanford University), Eric Xu (Stanford University), Atul Shenoy (Microsoft), Asvin Ananthanarayan (Microsoft), John Sheu (Microsoft), Erik Meijer (Facebook), Xi Wu (Google), Jeff Naughton (Google), Peter D Bailis (Stanford University), and Matei Zaharia (Stanford University and Databricks)*

**Res:16.3 - Explain3D: Explaining Disagreements in Disjoint Datasets**

*Authors: Xiaolan Wang (University of Massachusetts Amherst) and Alexandra Meliou (University of Massachusetts Amherst)*

**Res:16.4 - Exploring Change - A New Dimension of Data Analytics**

*Authors: Tobias Bleifuß (Hasso Plattner Institute), Leon Bornemann (Hasso Plattner Institute), Theodore Johnson (AT&T Labs - Research), Dmitri V. Kalashnikov (AT&T Labs Research), Felix Naumann (Hasso Plattner Institute), and Divesh Srivastava (AT&T Labs Research)*

**Res:16.5 - ProvCite: Provenance-based Data Citation**

*Authors: Yinjun Wu (University of Pennsylvania), Abdussalam Alawini (University of Illinois at Urbana-Champaign), Daniel Deutch (Tel Aviv University), Tova Milo (Tel Aviv University), and Susan Davidson (University of Pennsylvania)*

**16:00 - 17:30****Demo Session: Group B (See page 104 for details)****Room: Hollywood Ballroom****16:00 - 17:30****Tutorial 4: The Ever Evolving Online Labor Market: Overview, Challenges and Opportunities****Room: Avalon**

*Speakers:* Sihem Amer-Yahia (CNRS, Univ. Grenoble Alpes) and Senjuti Basu Roy (NJIT)

*Description:* The goal of this tutorial is to make the audience aware of various discipline-specific research activities that could be characterized to be part of online labor markets and advocate for a unified framework that is interdisciplinary in nature and requires convergence of different research disciplines. We will discuss how such a framework could bring transformative effect on the nexus of humans, technology, and the future of work. The slides for this tutorial can be downloaded at: <https://centers.njit.edu/bdal/node/69/>

**16:00 - 17:30****Invited Industry Talks 2: Data through the Looking Glass: Performance, In-Mem, & AI****Room: Sacramento / San Francisco***Chair: Pat Helland (Salesforce, USA)**Adrian Colyer (Accel)***Performance in the Spotlight**

**Abstract:** Performance in its various guises features prominently in research evaluations, and rightly so. Without adequate performance a system is not fit for purpose. That doesn't necessarily mean we should pursue performance at all costs though. In this talk we'll explore a variety of additional evaluation criteria, with a focus on those that are most important to practitioners, and ask whether or not considering them can open up interesting avenues of research.



**Short Bio:** Adrian is the author of "The Morning Paper", a popular computer science research blog. When he's not reading and summarising research papers, he also works as a Venture Partner for Accel, where it's his job to help find and build great technology companies out of Europe and Israel. Prior to joining Accel, Adrian held CTO roles at a variety of companies including SpringSource, VMware, and Pivotal.

*Alex Boehm (SAP SE)*

### **In-memory for the Masses: Enabling Cost-efficient Deployments of In-memory Data Management Platforms for Business Applications**

**Abstract:** With unrivaled performance, modern in-memory data management platforms such as SAP HANA enable the creation of novel types of business applications. By keeping all data in memory, applications may combine both demanding transactional as well as complex analytical workloads in the context of a single system. While this excellent performance, data freshness, and flexibility gain is highly desirable in a vast range of modern business applications, the corresponding large appetite for main memory has significant implications on server sizing. Particularly, hardware costs on premise as well as in the cloud are at risk to increase significantly, driven by the high amount of DRAM that needs to be provisioned potentially. In this talk, we discuss a variety of challenges and opportunities that arise when running business applications in a cost-efficient manner on in-memory database systems. We discuss design possibilities on the database level such as various forms of data compression and compact processing with corresponding trade-offs. We then illustrate how (modern) hardware can help to (automatically) offload parts of the data sets to cheaper storage media such as NVM, SSDs, and disk, without significantly compromising performance from an application point of view. Finally, we illustrate how applications may facilitate TCO-efficient database deployments by modifications to (physical) schema design, as well as synergies with orthogonal, but highly related application topics such as data archiving, as mandated by data privacy regulations.

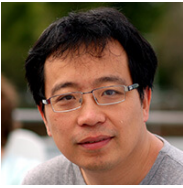


**Short Bio:** Dr. Alexander Boehm is one of the chief architects working on SAP's HANA in-memory database management system. His focus is on performance optimization and holistic improvements of enterprise architectures, in particular application server/DBMS co-design. Prior to joining SAP in 2010, he received his PhD from the University of Mannheim, Germany, where he worked on the development of efficient and scalable applications using declarative message processing.

*Haixun Wang (WeWork)*

### **Innovation Around Data and AI for the Physical World**

**Abstract:** AI and Machine Learning are making big strides in the cyberspace. Yet, there has been limited progress with AI in the physical world. This is mostly due to the challenges around infrastructure — from the need to collect and analyze data in the physical world to the need for taking physical actions. With over 400 buildings around the world, WeWork has a fleet of spaces ripe for experimenting how to blend the physical and the digital. At this scale, every decision from day-to-day ones, such as how to schedule room cleaning, to billion-dollar ones, such as how to source our next building and location, becomes a non-trivial data science problem. While most research and development around smart buildings have focused on energy saving, we see an opportunity to broaden this area to focus on improving the quality of life in workspaces and making businesses more efficient. We believe that intelligent environments will help make space more efficient and the addition of human insight will make for a more engaging experience.



**Short Bio:** Haixun Wang is VP of Engineering and Distinguished Scientist at WeWork. He is an IEEE fellow and Editor in Chief of the IEEE Data Engineering Bulletin. Before joining WeWork, he was a Director of Natural Language Processing at Amazon. From 2015 to 2017, he led the NLP organization in Facebook working on query and document understanding. From 2013 to 2015, he was with Google Research, working on natural language processing. From 2009 to 2013, he led research in knowledge bases, graph systems, and text processing at Microsoft Research Asia. He had been a research staff member at IBM T. J. Watson Research Center

from 2000 - 2009. He was Technical Assistant to Stuart Feldman (Vice President of Computer Science of IBM Research) from 2006 to 2007, and Technical Assistant to Mark Wegman (Head of Computer Science of IBM Research) from 2007 to 2009. He received the Ph.D. degree in Computer Science from the University of California, Los Angeles in 2000. He has published more than 200 research papers in international journals and conference proceedings. He served as PC chairs of many academic conferences, and he is on the editorial board of journals such as IEEE Transactions of Knowledge and Data Engineering (TKDE) and Journal of Computer Science and Technology (JCST). He won the best paper award in ICDE 2015, 10-year best paper award in ICDM 2013, and best paper award of ER 2009.

**17:30 - 19:00**

**Posters 1: Research track papers presented on Tuesday and Wednesday and all industrial track papers**

**Room: Jogging Track**

**19:30 - 23:45**

**VLDB Conference Banquet at the Westin Bonaventure**

## THURSDAY – August 29

**8:00 - 9:00**

**Morning Coffee**

**Room: California Foyer**

**9:00 - 10:30**

**Conference Awards**

**Room: Sacramento / San Francisco**

**10:30 - 11:00**

**Coffee Break (30 min)**

**Room: California Foyer**

**11:00 - 12:30**

**Research Session 17: Distributed Systems II**

**Room: Santa Anita**

**Chair: Philippe Bonnet (IT University Copenhagen, Denmark)**

**Res:17.1 - Megaphone: Latency-conscious state migration for distributed streaming dataflows**

**Authors:** Moritz Hoffmann (ETH Zurich), Andrea Lattuada (ETH Zurich), Frank McSherry (ETH Zurich), Vasiliki Kalavri (ETH Zurich), John Liagouris (ETH Zurich), and Timothy Roscoe (ETH Zurich)

**Res:17.2 - Minimizing Cost by Reducing Scaling Operations in Distributed Stream Processing**

Authors: Michael Borkowski (TU Wien), Christoph Hochreiner (TU Wien), and Stefan Schulte (TU Wien)

**Res:17.3 - Performance-Optimal Filtering: Bloom Overtakes Cuckoo at High Throughput**

Authors: Harald Lang (TU Munich), Thomas Neumann (TU Munich), Alfons Kemper (TU Munich), and Peter Boncz (Centrum Wiskunde & Informatica)

**Res:17.4 - Scaling-Up In-Memory Datalog Processing: Observations and Techniques**

Authors: Zhiwei Fan (University of Wisconsin-Madison), Jianqiao Zhu (University of Wisconsin-Madison), Zuyu Zhang (University of Wisconsin-Madison), Aws Albarghouthi (University of Wisconsin-Madison), Paraschos Koutris (University of Wisconsin-Madison), and Jignesh Patel (University of Wisconsin-Madison)

**Res:17.5 - Selectivity Estimation for Range Predicates using Lightweight Models**

Authors: Anshuman Dutt (Microsoft Research), Chi Wang (Microsoft Research), Azade Nazi (Google AI), Srikanth Kandula (Microsoft Research), Vivek Narasayya (Microsoft Research), and Surajit Chaudhuri (Microsoft Research)

11:00 - 12:30

**Research Session 18: Graph Analytics II****Room: San Jose****Chair: Xiangmin Zhou (RMIT University, Australia)****Res:18.1 - Constrained Shortest Path Query in a Large Time-Dependent Graph**

Authors: Ye Yuan (Beijing Institute of Technology), Xiang Lian (Kent State University), Guoren Wang (Beijing Institute of Technology), Yuliang Ma (Northeastern University), and Yishu Wang (Northeastern University)

**Res:18.2 - Correlation Constraint Shortest Path over Large Multi-Relation Graphs**

Authors: Xiaofei Zhang (University of Memphis) and Tamer Özsu (University of Waterloo)

**Res:18.3 - An In-Depth Comparison of s-t Reliability Algorithms over Uncertain Graphs**

Authors: Xiangyu Ke (Nanyang Technological University), Arijit Khan (Nanyang Technological University), and Leroy Lim (Nanyang Technological University)

**Res:18.4 - Motivo: Fast Motif Counting via Succinct Color Coding and Adaptive Sampling**

Authors: Marco Bressan (Sapienza University of Rome), Stefano Leucci (MPI-INF), and Alessandro Panconesi (Sapienza University of Rome)

**Res:18.5 - Querying Shortest Paths on Time Dependent Road Networks**

Authors: Yong Wang (Tsinghua University), Guoliang Li (Tsinghua University), and Nan Tang (Qatar Computing Research Institute, HBKU)

11:00 - 12:30

**Research Session 19: Indexing and Search****Room: San Diego****Chair: Bolin Ding (Alibaba Group, China)****Res:19.1 - Automatic Index Selection for Large-Scale Datalog Computation**

Authors: Pavle Subotic (UCL), Herbert Jordan (University of Innsbruck), Lijun Chang (The University of Sydney), Alan Fekete (University of Sydney), and Bernhard Scholz (University of Sydney)

**Res:19.2 - List Intersection for Web Search: Algorithms, Cost Models, and Optimizations**

Authors: Sunghwan Kim (POSTECH), Taesung Lee (IBM Research AI), Seungwon Hwang (Yonsei University), and Sameh Elnikety (Microsoft Research)

**Res:19.3 - Document Reordering for Faster Intersection**

*Authors: Qi Wang (New York University) and Torsten Suel (New York University)*

**Res:19.4 - Fast Approximate Nearest Neighbor Search With The Navigating Spreading-out Graph**

*Authors: Cong Fu (Zhejiang University), Chao Xiang (Zhejiang University), Changxu Wang (Zhejiang University), and Deng Cai (Zhejiang University)*

**Res:19.5 - PS-Tree-Based Efficient Boolean Expression Matching for High-Dimensional and Dense Workloads**

*Authors: Shuping Ji (Technical University of Munich) and Hans-Arno Jacobsen (University of Toronto)*

**11:00 - 12:30****Research Session 20: Usability and data mining****Room: Emerald Bay****Chair: Jun Yang (Duke University, USA)****Res:20.1 - User Guidance for Efficient Fact Checking**

*Authors: Thanh Tam Nguyen (EPFL), Matthias Weidlich (Humboldt-Universität zu Berlin), Hongzhi Yin (The University of Queensland), Bolong Zheng (Huazhong University of Science and Technology), Quoc Viet Hung Nguyen (Griffith University), and Bela Stantic (Griffith University)*

**Res:20.2 - Query Log Compression for Workload Analytics**

*Authors: Ting Xie (University at Buffalo, SUNY), Varun Chandola (University at Buffalo, SUNY), and Oliver A Kennedy (University at Buffalo, SUNY)*

**Res:20.3 - Subjective Databases**

*Authors: Yuliang Li (Megagon Labs), Aaron Feng (Megagon Labs), Jinfeng Li (Megagon Labs), Saran Mumick (Megagon Labs), Alon Halevy (Megagon Labs), Vivian Li (Megagon Labs), and Wang-Chiew Tan (Megagon Labs)*

**Res:20.4 - Example-Driven Query Intent Discovery: Abductive Reasoning using Semantic Similarity**

*Authors: Anna Fariha (University of Massachusetts Amherst) and Alexandra Meliou (University of Massachusetts Amherst)*

**11:00 - 12:30****Demo Session: Group A (See page 104 for details)****Room: Hollywood Ballroom****11:00 - 12:30****Tutorial 5: Machine Learning Meets Big Spatial Data****Room: Avalon**

**Speakers:** Ibrahim Sabek (University of Minnesota) and Mohamed F. Mokbel (Qatar Computing Research Institute, Hamad bin Khalifa University)

**Description:** The proliferation in amounts of generated data has propelled the rise of scalable machine learning solutions to efficiently analyze and extract useful insights from such data. Meanwhile, spatial data has become ubiquitous, e.g., GPS data, with increasingly sheer sizes in recent years. The applications of big spatial data span a wide spectrum of interests including tracking infectious disease, climate change simulation, drug addiction, among others. Consequently, major research efforts are exerted to support efficient analysis and intelligence inside these applications by either providing spatial extensions to existing machine learning solutions or building new solutions from scratch. In this 90-minutes tutorial, we comprehensively review the state-of-the-art work in the intersection of machine learning and big spatial data. We

cover existing research efforts and challenges in three major areas of machine learning, namely, data analysis, deep learning and statistical inference, as well as two advanced spatial machine learning tasks, namely, spatial features extraction and spatial sampling. We also highlight open problems and challenges for future research in this area. The slides for this tutorial can be download at <https://www-users.cs.umn.edu/~sabek001/v1ldb-2019-tutorial/>

**11:00 - 12:30**

### Industry Session 3: Implementing and Measuring High Performance Databases

**Room: Sacramento / San Francisco**

**Chair: Jingren Zhou (Alibaba Group, China)**

#### Ind:3.1 - S3: A Scalable In-memory Skip-List Index for Key-Value Store

*Authors: Jingtian Zhang (Zhejiang University), Sai Wu (Zhejiang University), Zeyuan Tan (Zhejiang University), Gang Chen (Zhejiang University), Zhushi Cheng (Alibaba Group), Wei Cao (Alibaba Group), Yusong Gao (Alibaba Group), and Xiaojie Feng (Alibaba Group)*

#### Ind:3.2 - Native Store Extension for SAP HANA

*Authors: Reza Sherkat (SAP SE), Colin Florendo (SAP SE), Mihnea Andrei (SAP SE), Rolando Blanco (SAP SE), Adrian Dragusanu (SAP SE), Amit Pathak (SAP SE), Pushkar Khadilkar (SAP SE), Neeraj Kulkarni (SAP SE), Christian Lemke (SAP SE), Sebastian Seifert (SAP SE), Sarika Iyer (SAP SE), Sasikanth Gottapu (SAP SE), Robert Schulze (SAP SE), Chaitanya Gottipati (SAP SE), Nirvik Basak (SAP SE), Yanhong Wang (SAP SE), Vivek Kandiyannallur (SAP SE), Santosh Pendap (SAP SE), Dheren Gala (SAP SE), Rajesh Almeida (SAP SE), and Prasanta Ghosh (SAP SE)*

#### Ind:3.3 - Choosing A Cloud DBMS: Architectures and Tradeoffs

*Authors: Junjay Tan (Brown University), Thanaa Ghanem (Metropolitan State University (Minnesota)), Matthew Perron (MIT CSAIL), Xiangyao Yu (MIT CSAIL), Michael Stonebraker (MIT CSAIL, Tamr, Inc.), David DeWitt (MIT CSAIL), Marco Serafini (University of Massachusetts Amherst), Ashraf Aboulnaga (Qatar Computing Research Institute, HBKU), and Tim Kraska (MIT CSAIL)*

#### Ind:3.4 - Adapting TPC-C Benchmark to Measure Performance of Multi-Document Transactions in MongoDB

*Authors: Asya Kamsky (MongoDB Inc)*

#### Ind:3.5 - A Morsel-Driven Query Execution Engine for Heterogeneous Multi-Cores

*Authors: Kayhan Dursun (Brown University), Carsten Binnig (TU Darmstadt), Ugur Cetintemel (Brown University), Garret Swart (Oracle Corporation), and Weiwei Gong (Oracle Corporation)*

**12:30 - 14:00**

**Lunch Break (90 min)**

**Room: Pasadena**

**14:00 - 15:30**

### Research Session 21: Privacy and Security

**Room: Santa Anita**

**Chair: Ashwin Machanavajjhala (Duke University, USA)**

#### Res:21.1 - A Comparative Evaluation of Order-Revealing Encryption Schemes and Secure Range-Query Protocols

*Authors: Dmytro Bogatov (Boston University), George Kollios (Boston University), and Leonid Reyzin (Boston University)*

#### Res:21.2 - Obscure: Information-Theoretic Oblivious and Verifiable Aggregation Queries

*Authors: Peeyush Gupta (UC Irvine), Yin Li (Xinyang Normal University), Sharad Mehrotra*



(UC Irvine), Nisha Panwar (UC Irvine), Shantanu Sharma (UC Irvine), and Sumaya Almanee (UC Irvine)

**Res:21.3 - Answering Range Queries Under Local Differential Privacy**

Authors: Graham Cormode (University of Warwick), Tejas Kulkarni (University Of Warwick), and Divesh Srivastava (AT&T Labs Research)

**Res:21.4 - PrivateSQL: A Differentially Private SQL Query Engine**

Authors: Ios Kotsogiannis (Duke University), Yuchao Tao (Duke University), Xi He (University of Waterloo), Maryam Fanaeepour (Duke University), Ashwin Machanavajhala (Duke University), Michael Hay (Colgate University), and Gerome Miklau (University of Massachusetts Amherst)

**Res:21.5 - Arx: An Encrypted Database using Semantically Secure Encryption**

Authors: Rishabh Poddar (UC Berkeley), Tobias Boelter (UC Berkeley), and Raluca Ada Popa (UC Berkeley)

14:00 - 15:30

**Research Session 22: Clustering and Outlier Detection**

**Room: San Jose**

**Chair: Ye Yuan (Beijing Institute of Technology, China)**

**Res:22.1 - Efficient Discovery of Sequence Outlier Patterns**

Authors: Lei Cao (MIT), Yizhou Yan (Worcester Polytechnic Institute), Samuel Madden (MIT), Elke Rundensteiner (Worcester Polytechnic Institute), and Mathan Gopalsamy (Signify Research, Cambridge, MA USA)

**Res:22.2 - NETS: Extremely Fast Outlier Detection from a Data Stream via Set-Based Processing**

Authors: Susik Yoon (KAIST), Jae-Gil Lee (KAIST), and Byung Suk Lee (University of Vermont)

**Res:22.3 - Solving k-center Clustering (with Outliers) in MapReduce and Streaming, almost as Accurately as Sequentially.**

Authors: Matteo Ceccarello (IT University of Copenhagen), Andrea Pietracaprina (University of Padova), and Geppino Pucci (University of Padova)

**Res:22.4 - Efficient and Effective Algorithms for Clustering Uncertain Graphs**

Authors: Kai Han (University of Science and Technology of China), Fei Gui (University of Science and Technology of China), Xiaokui Xiao (National University of Singapore), Jing Tang (National University of Singapore), Yuntian He (University of Science and Technology of China), Zongmai Cao (University of Science and Technology of China), and He Huang (Soochow University)

**Res:22.5 - k/2-hop: Fast Mining of Convoy Patterns With Effective Pruning**

Authors: Faisal Moeen Orakzai (ULB), Toon Calders (Universiteit Antwerpen), and Torben Bach Pedersen (Aalborg University)

14:00 - 15:30

**Research Session 23: Data Integration**

**Room: San Diego**

**Chair: Paolo Papotti (Eurecom, France)**

**Res:23.1 - Rewriting of Plain SO Tgds into Nested Tgds**

Authors: Rihan Hai (RWTH Aachen University) and Christoph Quix (Hochschule Niederrhein & Fraunhofer FIT)

**Res:23.2 - Embedded Functional Dependencies and Data-completeness Tailored Database Design**

*Authors: Ziheng Wei (University of Auckland) and Sebastian Link (University of Auckland)*

**Res:23.3 - Distributed Implementations of Dependency Discovery Algorithms**

*Authors: Hemant Saxena (University of Waterloo), Lukasz Golab (University of Waterloo), and Ihab F Ilyas (University of Waterloo)*

**Res:23.4 - Efficient Knowledge Graph Accuracy Evaluation**

*Authors: Junyang Gao (Duke University), Xian Li (Amazon.com), Yifan Ethan Xu (Amazon.com), Bunyamin Sisman (Amazon.com), Xin Luna Dong (Amazon.com), and Jun Yang (Duke University)*

**Res:23.5 - Ontology-based Entity Matching in Attributed Graphs**

*Authors: Hanchao Ma (Washington State University), Morteza Alipourlangouri (McMaster University, Canada), Yinghui Wu (Washington State University), Fei Chiang (McMaster University), and Jiaying Pi (Siemens Corporate Technology)*

**14:00 - 15:30**

**Research Session 24: Blockchain**

**Room: Emerald Bay**

**Chair: C. Mohan (IBM Research-Almaden, USA)**

**Res:24.1 - CAPER: A Cross-Application Permissioned Blockchain**

*Authors: Mohammad Javad Amiri (University of California, Santa Barbara), Divyakant Agrawal (University of California, Santa Barbara), and Amr El Abbadi (University of California, Santa Barbara)*

**Res:24.2 - (Best Paper) Fine-Grained, Secure and Efficient Data Provenance on Blockchain Systems**

*Authors: Pingcheng Ruan (National University of Singapore), Gang Chen (Zhejiang University), Tien Tuan Anh Dinh (National University of Singapore), Qian Lin (National University of Singapore), Beng Chin Ooi (National University of Singapore), and Meihui Zhang (Beijing Institute of Technology)*

**Res:24.3 - Blockchain Meets Database: Design and Implementation of a Blockchain Relational Database**

*Authors: Senthil Nathan (IBM Research - India), Chander Govindarajan (IBM Research - India), Adarsh Saraf (IBM Research - India), Manish Sethi (IBM Industry Platforms, USA), and Praveen Jayachandran (IBM Research - India)*

**Res:24.4 - BlockchainDB - A Shared Database on Blockchains**

*Authors: Muhammad El-Hindi (TU Darmstadt), Carsten Binnig (TU Darmstadt), Arvind Arasu (Microsoft), Donald Kossmann (Microsoft Research), and Ravi Ramamurthy (Microsoft)*

**14:00 - 15:30**

**Demo Session: Group B (See page 104 for details)**

**Room: Hollywood Ballroom**

**14:00 - 15:30**

**Tutorial 6: TextCube: Automated Construction and Multi-dimensional Exploration**

**Room: Avalon**

**Speakers:** Yu Meng (University of Illinois at Urbana-Champaign), Jiaxin Huang (University of Illinois Urbana-Champaign), Jingbo Shang (University of Illinois at Urbana-Champaign), and Jiawei Han (University of Illinois at Urbana-Champaign)

**Description:** Today's society is immersed in a wealth of text data, ranging from news articles, to social media, research literature, medical records, and corporate reports. A grand challenge of data science and engineering is to develop effective and scalable methods to extract

structures and knowledge from massive text data to satisfy diverse applications, without extensive, corpus-specific human annotations. In this tutorial, we show that TextCube provides a critical information organization structure that will satisfy such an information need. We overview a set of recently developed data-driven methods that facilitate automated construction of TextCubes from massive, domain-specific text corpora, and show that TextCubes so constructed will enhance text exploration and analysis for various applications. We focus on new TextCube construction methods that are scalable, weakly-supervised, domain-independent, language-agnostic, and effective (i.e., generating quality TextCubes from large corpora of various domains). We will demonstrate with real datasets (including news articles, scientific publications, and product reviews) on how TextCubes can be constructed to assist multidimensional analysis of massive text corpora. The slides for this tutorial can be downloaded at <https://shangjingbo1226.github.io/2019-08-23-vldb-tutorial/>.

**14:00 - 15:30**

**Industry Session 4: Pretty Good Answers and Pretty Darned Fast!**

**Room: Sacramento / San Francisco**

**Chair: Yuanyuan Tian (IBM Research-Almaden, USA)**

**Ind:4.1 - DDSketch: A Fast and Fully-Mergeable Quantile Sketch with Relative-Error Guarantees**

*Authors: Charles Masson (Datadog), Jee E. Rim (Datadog), and Homin K. Lee (Datadog)*

**Ind:4.2 - Experiences with Approximating Queries in Microsoft's Production Big-Data Clusters**

*Authors: Srikanth Kandula (Microsoft), Kukjin Lee (Microsoft), Surajit Chaudhuri (Microsoft), and Marc Friedman (Microsoft)*

**Ind:4.3 - SAP HANA goes private - From Privacy Research to Privacy Aware Enterprise Analytics**

*Authors: Stephan Kessler (SAP SE), Jens Hoff (SAP SE), and Johann-Christoph Freytag (Humboldt-Universität zu Berlin)*

**Ind:4.4 - Tunable Consistency in MongoDB**

*Authors: William Schultz (MongoDB, Inc.), Tess Avitabile (MongoDB, Inc.), and Alyson Cabral (MongoDB, Inc.)*

**Ind:4.5 - Customizable and Scalable Fuzzy Join for Big Data**

*Authors: Zhimin Chen (Microsoft Research), Yue Wang (Microsoft Research), Vivek Narasayya (Microsoft Research), and Surajit Chaudhuri (Microsoft Research)*

**15:30 - 16:00**

**Coffee Break (30 min)**

**Room: California Foyer**

**16:00 - 17:30**

**Research Session 25: Spatial Data**

**Room: Santa Anita**

**Chair: Vincent Oria (NJIT, USA)**

**Res:25.1 - Finding Attribute-Aware Similar Region for Data Analysis**

*Authors: Kaiyu Feng (Nanyang Technological University), Gao Cong (Nanyang Technological University), Christian S Jensen (Aalborg University), and Tao Guo (Google)*

**Res:25.2 - Ridesharing: Simulator, Benchmark, and Evaluation**

*Authors: James J Pan (Tsinghua University), Guoliang Li (Tsinghua University), and Juntao Hu (Beihang University)*

**Res:25.3 - The Maximum Trajectory Coverage Query in Spatial Databases**

*Authors: Mohammed Eunus Ali (Bangladesh University of Engineering and Technology (BUET)), Shadman Saqib Eusuf (Bangladesh University of Engineering and Technology (BUET)), Kaysar Abdullah (Bangladesh University of Engineering and Technology (BUET)), Farhana M. Choudhury (RMIT University), J. Shane Culpepper (RMIT University), and Timos Sellis (Swinburne University of Technology)*

**Res:25.4 - The Flexible Socio Spatial Group Queries**

*Authors: Bishwamitra Ghosh (Bangladesh University of Engineering and Technology (BUET)), Mohammed Eunus Ali (Bangladesh University of Engineering and Technology (BUET)), Farhana M Choudhury (RMIT University), Sajid Hasan (Bangladesh University of Engineering and Technology (BUET)), Timos Sellis (Swinburne University of Technology), and Jianxin Li (The University of Western Australia)*

**Res:25.5 - Creating Top Ranking Options in the Continuous Option and Preference Space**

*Authors: Bo Tang (Southern University of Science and Technology), Kyriakos Mouratidis (Singapore Management University), Man Lung Yiu (Hong Kong Polytechnic University), and Zhenyu Chen (Southern University of Science and Technology)*

**16:00 - 17:30****Research Session 26: Transaction Processing II****Room: San Jose****Chair: Yingjun Wu (IBM Research - Almaden)****Res:26.1 - IPA: Invariant-Preserving Applications for Weakly Consistent Replicated Databases**

*Authors: Valter Balegas (NOVA LINCS, FCT, Universidade NOVA de Lisboa), Sérgio Duarte (NOVA LINCS, FCT, Universidade NOVA de Lisboa), Carla Ferreira (NOVA LINCS, FCT, Universidade NOVA de Lisboa), Rodrigo Rodrigues (INESC-ID & Instituto Superior Técnico, Universidade de Lisboa), and Nuno Preguiça (NOVA LINCS, FCT, Universidade NOVA de Lisboa)*

**Res:26.2 - MgCrab: Transaction Crabbing for Live Migration in Deterministic Database Systems**

*Authors: Yu-Shan Lin (National Tsing Hua University), Shao-Kan Pi (National Tsing Hua University), Meng-Kai Liao (National Tsing Hua University), Ching Tsai (National Tsing Hua University), Aaron J Elmore (University of Chicago), and Shan-Hung Wu (National Tsing Hua University)*

**Res:26.3 - STAR: Scaling Transactions through Asymmetric Replication**

*Authors: Yi Lu (MIT), Xiangyao Yu (MIT), and Samuel Madden (MIT)*

**Res:26.4 - Adaptive Optimistic Concurrency Control for Heterogeneous Workloads**

*Authors: Jinwei Guo (East China Normal University), Peng Cai (East China Normal University), Jiahao Wang (East China Normal University), Weining Qian (East China Normal University), and Aoying Zhou (East China Normal University)*

**Res:26.5 - Improving Optimistic Concurrency Control Through Transaction Batching and Operation Reordering**

*Authors: Bailu Ding (Microsoft Research), Lucja Kot (GrammaTech, Inc), and Johannes Gehrke (Microsoft Corporation)*

**16:00 - 17:30****Research Session 27: Potpourri****Room: San Diego****Chair: Senjuti Basu Roy (NJIT, USA)****Res:27.1 - Efficient Task-Specific Data Valuation for Nearest Neighbor Algorithms**

*Authors: Ruoxi Jia (UC Berkeley), David Dao (ETH Zurich), Boxin Wang (Zhejiang University), Frances Ann Hubis (ETH Zurich), Nezihe Merve Gürel (ETH Zurich), Bo Li (University of Illinois at Urbana–Champaign), Ce Zhang (ETH Zurich), Costas J. Spanos (UC Berkeley), and Dawn Song (UC Berkeley)*

**Res:27.2 - An Experimental Evaluation of Garbage Collectors on Big Data Applications**

*Authors: Lijie Xu (Chinese Academy of Sciences), Tian Guo (Worcester Polytechnic Institute), Wensheng Dou (Chinese Academy of Sciences), Wei Wang (Chinese Academy of Sciences), and Jun Wei (Chinese Academy of Sciences)*

**Res:27.3 - A Comparative Evaluation of Systems for Scalable Linear Algebra-based Analytics**

*Authors: Anthony Thomas (University of California, San Diego) and Arun Kumar (University of California, San Diego)*

**Res:27.4 - GRAIL: Efficient Time-Series Representation Learning**

*Authors: John Paparrizos (University of Chicago) and Michael Franklin (University of Chicago)*

**Res:27.5 - ShrinkWrap: Efficient SQL Query Processing in Differentially Private Data Federations**

*Authors: Johes Bater (Northwestern University), Xi He (University of Waterloo), William Ehrich (Northwestern University), Ashwin Machanavajhala (Duke University), and Jennie Rogers (Northwestern University)*

**16:00 - 17:30**

**Research Session 28: Query Processing II**

**Room: Emerald Bay**

*Chair: Ilaria Bartolini (University of Bologna, Italy)*

**Res:28.1 - A Concave Path to Low-overhead Robust Query Processing**

*Authors: Srinivas Karthik (Indian Institute of Science), Jayant R. Haritsa (Indian Institute of Science), Sreyash Kenkre (IBM Research), and Vinayaka Pandit (IBM Research)*

**Res:28.2 - Interactive Summarization and Exploration of Top Aggregate Query Answers**

*Authors: Yuhao Wen (Duke University), Xiaodan Zhu (Duke University), Sudeepa Roy (Duke University), and Jun Yang (Duke University)*

**Res:28.3 - Durable Top-k Queries on Temporal Data**

*Authors: Junyang Gao (Duke University), Pankaj K. Agarwal (Duke University), and Jun Yang (Duke University)*

**Res:28.4 - Scalable, Variable-Length Similarity Search in Data Series: The ULISSE Approach**

*Authors: Michele Linardi (Paris Descartes University) and Themis Palpanas (Paris Descartes University)*

**Res:28.5 - Optimizing Subgraph Queries by Combining Binary and Worst-Case Optimal Joins**

*Authors: Amine Mhedhbi (University of Waterloo) and Semih Salihoglu (University of Waterloo)*

**16:00 - 17:30**

**Tutorial 6: TextCube: Automated Construction and Multi-dimensional Exploration**

**Room: Avalon**

*Speakers: Yu Meng (University of Illinois at Urbana-Champaign), Jiaxin Huang (University of Illinois Urbana-Champaign), Jingbo Shang (University of Illinois at Urbana-Champaign), and*

Jiawei Han (University of Illinois at Urbana-Champaign)

**Description:** Today's society is immersed in a wealth of text data, ranging from news articles, to social media, research literature, medical records, and corporate reports. A grand challenge of data science and engineering is to develop effective and scalable methods to extract structures and knowledge from massive text data to satisfy diverse applications, without extensive, corpus-specific human annotations. In this tutorial, we show that TextCube provides a critical information organization structure that will satisfy such an information need. We overview a set of recently developed data-driven methods that facilitate automated construction of TextCubes from massive, domain-specific text corpora, and show that TextCubes so constructed will enhance text exploration and analysis for various applications. We focus on new TextCube construction methods that are scalable, weakly-supervised, domain-independent, language-agnostic, and effective (i.e., generating quality TextCubes from large corpora of various domains). We will demonstrate with real datasets (including news articles, scientific publications, and product reviews) on how TextCubes can be constructed to assist multidimensional analysis of massive text corpora. The slides for this tutorial can be downloaded at <https://shangjingbo1226.github.io/2019-08-23-vldb-tutorial/>.

16:00 - 17:30

**Tutorial 7: Personal Database Security and Trusted Execution Environments: A Tutorial at the Crossroads**

**Room: Sacramento / San Francisco**

**Speakers:** Nicolas AnCIAUX (Inria Saclay, U. Versailles Saint-Quentin, Université Paris-Saclay), Luc Bouganim (Inria Saclay, U. Versailles Saint-Quentin, Université Paris-Saclay), Philippe Pucheral (U. Versailles Saint-Quentin, Inria Saclay, Université Paris-Saclay), Iulian Sandu Popa (U. Versailles Saint-Quentin, Inria Saclay, Université Paris-Saclay), and Guillaume Scerri (U. Versailles Saint-Quentin, Inria Saclay, Université Paris-Saclay)

**Description:** Smart disclosure initiatives and new regulations such as GDPR in the EU increase the interest for Personal Data Management Systems (PDMS) being provided to individuals to preserve their entire digital life. Consequently, the thorny issue of data security becomes more and more prominent, but highly differs from traditional privacy issues in outsourced corporate databases. Concurrently, the emergence of Trusted Execution Environments (TEE) changes the game in privacy-preserving data management with novel security models. This tutorial offers a global perspective of the current state of work at the confluence of these two rapidly growing areas. The goal is threefold: (1) review and categorize PDMS solutions and identify existing privacy threats and countermeasures; (2) review new security models capitalizing on TEEs and related privacy-preserving data management solutions relevant to the personal context; (3) discuss new challenges at the intersection of PDMS security and TEE-based data management. The slides for this tutorial can be downloaded at <https://team.inria.fr/petrus/TutorialVLDB2019/>.

17:30 - 18:30

**Posters 2: Research track papers presented on Thursday and VLDB Journal papers**

**Room: Jogging Track**

**FRIDAY – August 30 (WORKSHOPS)****9:00 - 17:30****BlockChain and Distributed Ledger (BCDL)****Room: Santa Anita***Chair: C. Mohan (IBM) and Beng Chin Ooi (National University of Singapore)***9:00 - 17:30****Big Data Open Source Systems (BOSS)****Room: Emerald Bay / Santa Monica***Chair: Pinar Tözün (IT University of Copenhagen) and Emanuel Zraggen (MIT)***9:00 - 17:30****Conversational Access to daTa (CAST)****Room: Mt. Washington***Chair: H.V Jagadish (University of Michigan), Georgia Koutrika (Athena Research Center), and Fatma Özcan (IBM Research - Almaden)***9:00 - 17:30****Data Management and Analytics for Medicine and Healthcare (DMAH)****Room: San Jose***Chair: Fusheng Wang (Stony Brook University), Gang Luo (University of Washington), Yanhui Liang (Google Inc), and Alevtina Dubovitskaya (Lucerne University of Applied Sciences and Arts and Swisscom)***9:00 - 17:30****Polystores including Managing Privacy, Security and Policy for Heterogenous Data (POLY)****Room: San Diego***Chair: Vijay Gadepally (MIT), Tim Mattson (Intel), Mike Stonebraker (MIT)***9:00 - 17:40****Very Large Internet of Things (VLIoT)****Room: Avalon***Chair: Sven Groppe (University of Lübeck, Germany) and Markus Endler (PUC Rio, Brasil)*

# **DETAILED WORKSHOP PROGRAM**



**MONDAY – August 26 (Emerald Bay)****ACCELERATING ANALYTICS AND DATA MANAGEMENT SYSTEMS USING MODERN PROCESSOR AND STORAGE ARCHITECTURES WORKSHOP (ADMS)**

**Chairs:** Rajesh Bordawekar (IBM T.J. Watson Research Center) and Tirthankar Lahiri (Oracle)

Please see workshop website for the most up to date schedule <http://www.adms-conf.org/>

**9:00 - 10:30****Session 1****Experimental Study of Memory Allocation for High-Performance Query Processing**

*Authors: Dominik Durner, TUM, Viktor Leis, Friedrich-Schiller-Universität Jena and Thomas Neumann, TUM.*

*Keynote: Pradeep Dubey (Intel)*

**AI: What Makes it Hard and Fun!**

**Abstract:** The confluence of massive data with massive compute is unprecedented. This coupled with recent algorithmic breakthroughs, we are now at the cusp of a major transformation. This transformation has the potential to disrupt a long-held balance between humans and machine where all forms of number crunching is left to computers, and most forms of decision-making is left to us humans. This transformation is spurring a virtuous cycle of compute which will impact not just how we do computing, but what computing can do for us. In this talk, I will discuss some of the application-level opportunities and system-level challenges that lie at the heart of this intersection of traditional high-performance computing with emerging data-intensive computing.

**10:30 - 11:00****Coffee Break****11:00 - 12:30****Session 2****GPU Accelerated Top-K Selection With Efficient Early Stopping**

*Authors: Vasileios Zois, Vassilis J. Tsotras, Wallid A. Najjar. University of California, Riverside*

**A Study on Database Cracking with GPUs**

*Authors: Eleazar Leal, University of Minnesota, and Le Gruenwald, University of Oklahoma*

**Efficient Quadtree Construction for Indexing Large-Scale Point Data on GPUs: Bottom-Up vs. Top-Down**

*Authors: Jianting Zhang, City University of New York, and Le Gruenwald, University of Oklahoma*

**Accelerating Regular Path Queries using FPGA**

*Authors: Kento Miura, Department of Computer Science University of Tsukuba, Toshiyuki Amagasa and Hiroyuki Kitagawa, Center for Computational Sciences University of Tsukuba*

**12:30 - 14:00****Lunch Break****14:00 - 15:30****Session 3****Computational Storage For Big Data Analytics**

*Authors: Balavinayagam Samynathan, Keith Chapman, Mehdi Nik, Behnam Robotmili, Shahrzad Mirkhani and Maysam Lavasani. Bigstream*

**Keynote: Berni Schiefer (Amazon Web Services)**

**Challenges and Opportunities for Acceleration in a Cloud-Native Data Warehouse**

**Abstract:** In this talk we will take a fresh look at techniques that can be used to accelerate ETL/ELT and Query Processing in a data warehouse. There are both unique opportunities, but also special challenges, for a Cloud-Native Data Warehouse. We will use a Cloud-Native Data Warehouse, Amazon Redshift, as our working example to illustrate what needs acceleration, what hardware and software techniques might apply and what unique opportunities and challenges exist for a Cloud-Native Data Warehouse.

**15:30 - 16:00****Coffee Break****16:00-17:30****Session 4****High-Performance In-Network Data Processing**

*Authors: Jaco Hofmann, Lasse Thostrup, Tobias Ziegler, Carsten Binnig and Andreas Koch.TU Darmstadt*

**Keynote: GPU Odyssey**

**Nikolay Sakharnykh (Nvidia)**

**Abstract:** Today's GPUs are no longer just video accelerators from 20 years ago crunching pixels and running a static graphics pipeline. They are complex "mini" supercomputers with lots of diverse high-throughput computational cores used to accelerate critical computational blocks in ray tracing, deep learning, and HPC workloads. The GPU programming models and tools are constantly evolving enabling developers to use the new capabilities and more efficiently utilize the hardware. NVIDIA RTX Technology provides a simple, recursive, and flexible pipeline for accelerating ray tracing algorithms, inspiring developers to explore new RTX applications and take advantage of the modern GPU architecture. High throughput computations demand high memory bandwidth. GPUs are pushing the limits for memory bandwidth approaching a terabyte per second rate, which makes them ideal for accelerating data analytics workloads. Core database operations, such as joins and aggregations, map naturally to the GPU architecture and, coupled with fast compression and NVLINK interconnect, enable running the most complex queries on the GPU, not possible before.

**MONDAY – August 26 (San Jose)****APPLIED AI FOR DATABASE SYSTEMS AND APPLICATIONS (AIDB)**

**Chairs:** Berthold Reinwald (IBM Almaden Research Center), Bingsheng He (National University of Singapore) and Yingjun Wu (IBM Almaden Research Center)

Please see workshop website for the most up to date schedule <https://sites.google.com/view/aidb2019/>

**9:00****Opening Remarks**

**Keynote:** *Andrew Pavlo (CMU)*

**9:00 - 9:45****Self-Driving Databases: The Hard Parts**

**Abstract:** The current research trend is on developing "learned" components to supplement and replace legacy components in database management systems (DBMSs). Such learned components use machine learning (ML) methods to identify non-trivial trends and correlations in the DBMS's runtime behavior. They then use this information to create execution strategies and data structures that are tailored to the application's access patterns. The hope is that learned components will enable new optimizations that are not possible today because the complexity of managing DBMSs has surpassed the abilities of humans. This could then lead to the ultimate goal of achieving a "self-driving" DBMS that is able to configure, manage, and optimize itself automatically as the database and its workload evolve over time. The bad news is that creating such a fully autonomous DBMS is harder than that. The problem requires both holistic systems engineering and novel ML solutions that cannot be solved with just adding learned components to an existing DBMS.

In this talk, I discuss the pressing unsolved problems in self-driving DBMSs. These include how to support training data collection, fast state changes, succinct state and action representations, and accurate reward observations. I will also present techniques on how to build a new autonomous DBMS or the steps needed to retrofit an existing one to enable automated control. This talk is part of the "My Wife is Pregnant" Speaking Tour. More information is available at <https://cmudb.io/tour2019>

**Bio:** Andy Pavlo is an Associate Professor of Databaseology in the Computer Science Department at Carnegie Mellon University.

**9:45 - 10:30****Session 1****Estimating Filtered Group-By Queries is Hard: Deep Learning to the Rescue.**

*Authors: Andreas Kipf (TUM), Michael J Freitag (TUM), Dimitri Vorona (TUM), Peter Boncz (Centrum Wiskunde & Informatica), Thomas Neumann (TUM), Alfons Kemper (TUM)*

**Accelerating B+tree Search by Using Simple Machine Learning Techniques.**

*Authors: Anisa LLaveshi (TUM), Utku Sirin (EPFL), Robert West (EPFL), Anastasia Ailamaki (EPFL)*

**10:30 - 11:00****Coffee Break**

Keynote: *Feifei Li (Alibaba & University of Utah)*

**Journey to SDDP: Towards Building a Self-Driving Database Platform**

**11:00 - 11:45**

**Abstract:** Self-Driving Database Platform (SDDP) provides cloud databases with automatic operation and maintenance, and enables intelligent database kernels (such as adaptive hot/cold data separation). The DBMS running on SDDP has the ability of self-detection, self-recovery, self-decision-making, self-optimization, and serving cloud users with transparent non-stop services. For cloud databases, tuning the buffer size appropriately is critical to the performance, since memory is usually the resource bottleneck. For large-scale databases supporting heterogeneous applications, configuring the individual buffer sizes for a significant number of database instances presents a scalability challenge. Manual optimization is neither efficient nor effective, and even not feasible for large cloud clusters, especially when the workload may dynamically change on each instance. The difficulty lies in the fact that each database instance requires a different buffer size that is highly individualized, subject to the constraint of the total buffer memory space. It is imperative to resort to algorithms that automatically orchestrate the buffer tuning for the entire database instances.

To this end, we enable SDDP to automatically tune buffer sizes based on a deep learning method, called iBTune. It has been deployed for more than 10,000 OLTP database instances in Alibaba production system. Specifically, it leverages the information from similar workloads to find out the tolerable miss ratio of each instance. Then, it utilizes the relationship between miss ratios and allocated memory sizes to individually optimize the target buffer sizes. To provide a guaranteed level of service level agreement (SLA), we design a pairwise deep neural network that uses features from measurements on pairs of instances to predict the upper bounds of the request response times. A target buffer size can be adjusted only when the predicted response time upper bound is in a safe limit. The successful deployment on Alibaba production environment, which safely reduces the memory footprint by more than 17% compared to the original system that relies on manual configurations, demonstrates the effectiveness of our solution.

**Bio:** Feifei Li is currently a Vice President of Alibaba Group, ACM Distinguished Scientist, President of the Database Products Business Unit of Alibaba Cloud and the Database and Storage Lab of DAMO academy. Before joining Alibaba, He has been a tenured full professor at the School of Computing, University of Utah. He has won multiple awards from NSF, ACM, IEEE, Visa, Google, HP, and Huawei. In particular, he is a recipient of the IEEE ICDE 2014 10 Years Most Influential Paper Award, ACM SIGMOD 2016 Best Paper Award, ACM SIGMOD 2015 Best System Demonstration Award, IEEE ICDE 2004 Best Paper Award, NSF Career Award by US National Science Foundation. He has been an associate editor, co-chairs, and core committee members for many prestigious journals and conferences.

**11:45 - 12:30**

**Session 2**

**Demonstrating Semantic SQL Queries over Relational Data using the AI-Powered Database.**

*Authors: Jose Neves (IBM Systems), Rajesh Bordawekar (IBM T. J. Watson Research Center), Elpida Tzortzatos (IBM Systems)*

**“Amnesia” - Towards Machine Learning Models That Can Forget User Data Very Fast.**

*Authors: Sebastian Schelter (New York University)*

**12:30 - 14:00**

**Lunch Break**

**14:00 - 14:45**      **Keynote: Johannes Gehrke (Microsoft)**  
**AI-Powered DBs and the Future of Software**

**Abstract:** Our industry is in a big revolution where we are replacing traditional, human-written software, with machine-learned models. This is also happening in the database field where we have initial results on machine-learned query optimizers, indices, and selectivity estimation. I will talk about this new generation of AI-powered database systems and how the practice of software development is changing as a result the AI shift.

**Bio:** Johannes Gehrke is a Technical Fellow at Microsoft in the Experiences and Devices Group, working on machine learning and Big Data. From 1999 to 2015, he was on the faculty in the Department of Computer Science at Cornell University where he graduated 25 PhD students. Johannes has received an NSF Career Award, a Sloan Research Fellowship, a Humboldt Research Award, the 2011 IEEE Computer Society Technical Achievement Award, and he is an ACM Fellow. He co-authored the undergraduate textbook “Database Management Systems (McGrawHill (2002),” currently in its third edition), and he was Program co-Chair of ACM KDD 2004, VLDB 2007, IEEE ICDE 2012, ACM SOCC 2014, and IEEE ICDE 2015.

**14:45 - 15:30**      **Session 3**

**Model-based Approximate Query Processing.**

*Authors: Moritz Kulessa (TU Darmstadt), Benjamin Hilprecht (TU Darmstadt), Alejandro Molina (TU Darmstadt), Carsten Binnig (TU Darmstadt), Kristian Kersting (TU Darmstadt)*

**3 Lessons Learned from Implementing a Deep Reinforcement Learning Framework for Data Exploration.**

*Authors: Amit Somech (Tel Aviv University), Tova Milo (Tel Aviv University), Ori Barel (Tel Aviv University)*

**15:30 - 16:00**      **Coffee Break**

**16:00 - 17:10**      **Innovative talks - 17.5 min each**

**Query Optimization and Deep Learning: Derailing the Hype Train.**

*Speakers: Ryan Marcus (MIT)*

**Learned Sketches for Cardinality Estimation.**

*Speakers: Yao Lu (Microsoft Research)*

**How do Database Technologies Help Machine Learning Systems?**

*Speakers: Zeke Wang (ETH Zurich)*

**ThunderML: Machine Learning Systems on Heterogeneous Architectures.**

*Speakers: Zeyi Wen (National University of Singapore)*

**17:10 - 17:30**      **Lightning talks - 10 min each**

**Learned Operator Cost Models.**

*Speakers: Jan Kossmann (Hasso Plattner Institute)*

**Learning from Query vs Learning from Data.**

*Speakers: Chenggang Wu (UC Berkeley)*

**MONDAY – August 26 (San Diego)****REAL-TIME BUSINESS INTELLIGENCE AND ANALYTICS (BIRTE)**

**Chairs:** Malu Castellanos (Teradata), Badrish Chandramouli (Microsoft Research), and Shimin Chen (Chinese Academy of Sciences)

Please see workshop website for the most up to date schedule <https://birte-conf.github.io/birte2019/index.html>

**9:00 - 10:30****Session 1: Opening and Research and Position papers****Opening Remarks**

Malu Castellanos (Teradata), Badrish Chandramouli (Microsoft)

**FAST : Fragment Assisted Storage for Efficient Query Execution in Read-only Databases (Research Paper)**

*Authors:* Vivek Hamirwasia (IIIT Hyderabad), Kamalakar Karlapalem (IIIT Hyderabad), Satyanarayana Valluri (Oracle)

**Towards a Real-time Unsupervised Estimation of Predictive Model Degradation (Position Paper)**

*Authors:* Tania Cerquitelli, Stefano Proto, Francesco Ventura, Daniele Apiletti, Elena Baralis (Politecnico di Torino)

**Scaling Ordered Stream Processing on Shared-Memory Multicores (Research Paper)**

*Authors:* Guna Prasaad (University of Washington), G. Ramalingam (Microsoft Research), Kaushik Rajan (Microsoft Research)

**Distributed Classification of Text Streams: Limitations, Challenges, and Solutions (Position Paper)**

*Authors:* Artem Trofimov (JetBrains Research, Saint Petersburg State University), Nikita Sokolov (ITMO University), Mikhail Shavkunov (National Research University Higher School of Economics), Igor E Kuralenok (JetBrains Research), Boris Novikov (National Research University Higher School of Economics)

**10:30 - 11:00****Coffee Break****11:00 - 12:40****Session 2: Keynote 1 and Demo papers**

**Keynote:** *Phil Bernstein (Microsoft Research)*

**Stream Processing in an Actor-Oriented Database System**

**Abstract:** Many of today's near-real-time, server applications are stateful and processor-intensive. Since their workload is often highly variable, they must be scalable and elastic. Example applications include multi-player games, social networking, mobile computing, telemetry,

and Internet of Things. These applications have stateless front-end processors that feed stateful objects in the cloud. Usually these objects are backed by key-value cloud storage, rather than by stored procedures in a database system. This enables the system to scale elastically by adding or removing inexpensive middle-tier servers. When the objects are single-threaded and do not share memory, they are called actors. There are dozens of programming frameworks for building actor applications, such as Akka, Erlang, Orbit, and Orleans. Although these applications do not use a database system, they can benefit from database abstractions, such as transactions, indexing, queries, replication, geo-distribution, and stream processing. We therefore propose a new type of database system, called an actor-oriented database system (AODB), which supports these features over middle-tier stateful objects. As an example feature, we will discuss how to add stream processing to an AODB. The AODB must support different stream processing engines and message transports as plug-ins. It must allow a mix of declarative queries and custom application code and enable fault-tolerant processing of events. We will explain how these requirements were met in the stream processing framework of Orleans.

**Bio:** Philip A. Bernstein is a Distinguished Scientist at Microsoft Research. He has published over 150 papers and two books on the theory and implementation of database systems, especially on transaction processing and data integration, which are still the major focus of his research. He is an ACM Fellow, a AAAS Fellow, a winner of ACM SIGMOD's Codd Innovations Award, a member of the Washington State Academy of Sciences, and the U.S. National Academy of Engineering. He received a B.S. degree from Cornell and M.Sc. and Ph.D. from University of Toronto.

#### **Demo 1: FASTER State Management for Timely Dataflow**

*Authors: Matthew J Brookes, Vasiliki Kalavri, John Liagouris (ETH Zurich)*

#### **Demo 2: Arcon: Continuous and Deep Data Stream Analytics**

*Authors: Paris Carbone, Max Meldrum (RISE Research Institutes of Sweden); Klas R Segeljakt, Lars Kroll, Christian Schulte, Seif Haridi (KTH Royal Institute of Technology)*

**12:40 - 14:00**

**Lunch Break**

**14:00 - 15:30**

**Session 3: Invited Industrial Talks**

#### **Apache Kafka and the Rise of Event-Driven Microservices**

*Authors: Jun Rao (Confluent)*

#### **The Materialize Incremental View Maintenance Engine**

*Authors: Frank McSherry (Materialize, Inc)*

**15:30 - 16:00**

**Coffee Break**

**16:00 - 18:00****Session 4: Keynote 2 and Panel****Keynote: Roger Barga (Amazon)  
Extracting Real-Time Insights from Streaming Data**

**Abstract:** Stream data processing is about identifying and responding to events happening in your business, in your service or application, or with your customers in near real-time. Sensors, IoT and mobile devices, and online transactions all generate data that can be monitored constantly to enable a business to detect and then act on events and insights before they lose their value. The need for large scale, real-time stream processing of big data in motion is more evident now than ever before. In this talk I will draw upon our experience with Amazon Kinesis data streaming services to highlight use cases and dive deep into the role of machine learning over streaming data to extract insights in real-time.

**Bio:** Roger Barga is General Manager for AWS Robotics and Autonomous Systems. Prior to that, Roger was General Manager and Director of Development for Amazon's Kinesis data streaming services. Before joining AWS, Roger was in the Cloud Machine Learning Group at Microsoft, where he was responsible for product management of the Azure Machine Learning service. Roger is also an affiliate professor at the University of Washington, where he is a lecturer in the Data Science and Machine Learning programs. Roger holds a PhD in computer science, has been granted over 30 patents, has published over 100 peer-reviewed technical papers and book chapters, and has authored a book on predictive analytics.

**Panel: IoT: The future of Real-Time Business Intelligence and Analytics**

*Moderator: Cyrus Shahabi (USC)*

**Participants:** Gustavo Alonso (ETH Zurich), Bhaskar Krishnamachari (USC), C. Mohan (IBM), Mohamed Sarwat (Arizona State University), Dave Schrader (Teradata)



**MONDAY – August 26 (Santa Anita)****PHD WORKSHOP**

**Chairs:** Ilaria Bartolini (University of Bologna) and Feifei Li (Alibaba)

Please see workshop website for the most up to date schedule <https://vladb.org/2019/?program-schedule-phd-workshop>

**9:00 - 10:30**

**Keynote: Johannes Gehrke (Microsoft)**  
**Database Systems 2.0**

**Abstract:** Software 2.0 – the augmentation and replacement of traditional code with ML models, especially deep neural networks – is changing how we develop, deploy, and maintain software. In this talk, I will describe this transition and its associated challenges for the database community. I will conclude with a set of open research challenges.

**Bio:** Johannes Gehrke is a Technical Fellow at Microsoft. From 1999 to 2015, he was on the faculty in the Department of Computer Science at Cornell University where he advised 25 PhD students. He co-authored the undergraduate textbook ‘Database Management Systems’ (McGrawHill, 2002), currently in its third edition. Johannes received the 2011 IEEE Computer Society Technical Achievement Award, and he is an ACM Fellow.

**10:30 - 11:00****Coffee Break****11:00 - 12:15****Technical Session 1: Big Data and Analytics (15 min each)**

**Efficient Scale-Out Using Query-Driven Workload Distribution and Fragment Allocation**  
*Authors: Stefan Halfpap (University of Potsdam)*

**Adopting Markov Logic Networks for Big Spatial Data and Applications**  
*Authors: Ibrahim Sabek (University of Minnesota)*

**Developing Big Data Analytics Architecture for Spatial Data**  
*Authors: Purnima Shah (Ahmedabad University)*

**Effective and Efficient Variable-Length Data Series Analytics**  
*Authors: Michele ML Linardi (Université de Paris)*

**From Distributed Sources to Distributed Sinks: Towards Truly Decentralized Event Stream Processing**  
*Authors: Samira Akili (Humboldt-University of Berlin)*

**12:15 - 14:00****Lunch Break****14:00-15:30**

**Keynote: Cong Yu (Google)**  
**Structured Data Meets News**

**Abstract:** The news ecosystem is going through profound changes that will have long-lasting impact on our civic society. One of the key challenges in the news ecosystem is how to encourage content consumption that will bring informational values to the users instead of purely consuming users' attention via any (i.e., unhealthy) means necessary. As part of the talk, I will describe how structured data can play an important role in helping users consume content in a more healthy way and present my long term vision on how our research community can contribute to this important cause. In the rest of the talk, I will describe various technical efforts, within the Structured Data Research Group at Google AI and in partnership with many product teams, on improving news consumption at Google. One specific example is the question and answering summarization of news articles, where we combine structured data, machine learning, and natural language processing techniques to help users understand news articles quickly.

**Bio:** Cong Yu is a research scientist and manager at Google Research in New York City and leads the Structured Data Research Group. The group's mission is to understand and leverage structured data on the Web to enhance user experience for Google products and has been responsible for several impactful products such as Web Tables, Structured Snippets, and Fact Checking at Google. Currently, his group focuses on technical research for news and has been partnering with journalists and policy advisors to combat online misinformation and polarization and to improve news consumption. His research interests are structured data exploration and mining, computational journalism, machine learning, natural language processing, and scalable data analysis. He was a keynote speaker for VLDB 2019 and twice served as an industrial program co-chair for VLDB (2014 and 2018). Outside of Google, he periodically teaches at NYU Courant's Department of Computer Science. Before Google, Cong was a Research Scientist at Yahoo! Research, also in NYC. He has a PhD from the University of Michigan, Ann Arbor, advised by Prof. H.V. Jagadish.

**15:30 - 16:00**

**Coffee Break**

**16:00 - 17:15**

**Technical Session 2: Unconventional Data (15 min each)**

**Pre-Estimation of Electric Vehicle Energy Consumption on Unfamiliar Roads and Actual Driving Experiments**

*Authors: Toshiaki Uemura (Yokohama National University)*

**Biomedical Data Categorization and Integration using Human-in-the-loop Approach**

*Authors: Priya Deshpande (DePaul University)*

**Provenance-Based Routing in Probabilistic Graph Databases**

*Authors: Yann Ramusat (ENS, CNRS, PSL University & Inria)*

**Make Informed Decisions: Understanding Query Results from Incomplete Databases**

*Authors: Poonam Kumari (State University of New York at Buffalo)*

**Optimized Spatio-Temporal Data Structures for Hybrid Transactional and Analytical Workloads on Columnar In-Memory Databases**

*Authors: Keven Richly (University of Potsdam)*

**17:15 - 18:30**

**Technical Session 3: Potpourri (15 min each)**

**Truly Scalable Data Series Similarity Search**

*Authors: Karima Echihabi (ENSIAS, Université Mohammed V de Rabat)*

**Data as a Language: A Novel Approach to Data Integration**

*Authors: Christos Koutras (Delft University of Technology)*

**Comprehensive Framework for Sorting Benchmarks**

*Authors: Sergey Madaminov (Stony Brook University)*

**A Cross-domain Natural Language Interface to Databases using Adversarial Text Method**

*Authors: Wenlu Wang (Auburn University)*

**Computational Aspects Around Preference Queries**

*Authors: Karim Alami (University of Bordeaux, CNRS)*

**MONDAY – August 26 (Avalon)****TECHNOLOGY CONFERENCE ON PERFORMANCE EVALUATION AND BENCHMARKING (TPCTC)**

**Chairs:** Raghunath Nambiar (AMD) and Meikel Poess (Oracle)

Please see workshop website for the most up to date schedule <http://www.tpc.org/tpctc/tpctc2019/>

**8:30 - 9:00**

**Opening Remarks, Welcome. The New TPC (Raghunath Nambiar)**

**9:00-10:30**

**Session 1: (30 min each)**

**Towards Benchmarking Cloud Big Data Services under SLA constraints**

*Authors: Nicolas Poggi, Victor Cuevas-Vicentín, David Carrera, Josep Lluís Berral, Thomas Fenech, Gonzalo Gomez, Davide Brini, Alejandro Montero, Umar Farooq Minhas, Jose A. Blakeley, Donald Kossmann, Raghu Ramakrishnan and Clemens Szyperski*

**Efficient Multiway Hash Join on Reconfigurable Hardware**

*Authors: Rekha Singhal, Yaqiz Zhang, Jeffrey D. Ullman, Raghu Prabhakar and Kunle Olukotun.*

**Challenges in Distributed MLPerf**

*Authors: Miro Hodak and Ajay Dholakia.*

**10:30-11:00**

**Coffee Break**

**11:00-12:30**

**Session 2: (30 min each)**

**AdaBench - Towards an Industry Standard Benchmark for Advanced Analytics**

*Authors: Tilmann Rabl, Christoph Brücke-Wendorff, Philipp Härtling, Stella Stars, Rodrigo Escobar Palacios, Hamesh Patel, Satyam Srivastava, Christoph Boden, Jens Meiners and Sebastian Schelter.*

**TPCxBB (BigBench) in a single node environment**

*Authors: Dippy Aggarwal, Shreyas Shekhar, Chris Elford, Umachandar Jayachandran, Sadashivan Krishnamurthy, Brendan Niebruegge and Jamie Reding.*

**CBench-Dynamo: A Consistency Benchmark for NoSQL Database Systems**

*Authors: Miguel Diogo, Bruno Cabral and Jorge Bernardino.*

**12:30 - 13:30**

**Lunch Break**

**Keynote: State of Permissionless and Permissioned Blockchains: Myths and Reality**

**13:30 - 14:30**

**C. Mohan (IBM Almaden Research Center)**

**Abstract:** It has been a decade since the concept of blockchain was invented as the underlying core data structure of the permissionless or public Bitcoin cryptocurrency network. Since then, several cryptocurrencies, and associated concepts like tokens and ICOs have emerged. After much speculation and hype, significant number of them have become problematic or worthless, even though some countries have embraced them! The permissionless blockchain system Ethereum emerged by generalizing the use of blockchains to manage any kind of asset, be it physical or purely digital, with the introduction of the concept of Smart Contracts. Over the years, numerous myths have developed with respect to the purported utility and the need for permissionless blockchains. The adoption and further adaptation of blockchains and smart contracts for use in the permissioned or private environments is what I consider to be useful and of practical consequence. Hence, the technical aspects of only private blockchain systems will be the focus of my TPCTC 2019 keynote. Along the way, I will bust many myths associated with permissionless blockchains. I will also compare traditional database technologies with blockchain systems' features and identify desirable future research topics. Extensive blockchain related collateral can be found at <http://bit.ly/CMbcDB>

**Bio:** Dr. C. Mohan is currently an IBM Fellow at the IBM Almaden Research Center in Silicon Valley and a Distinguished Visiting Professor at Tsinghua University in China. He has been an IBM researcher for 37 years in the database and related areas, impacting numerous IBM and non-IBM products, the research and academic communities, and standards, especially with his invention of the well-known ARIES family of database locking and recovery algorithms, and the Presumed Abort distributed commit protocol. This IBM (1997), ACM (2002) and IEEE (2002) Fellow has also served as the IBM India Chief Scientist (2006-2009). In addition to receiving the ACM SIGMOD Edgar F. Codd Innovations Award (1996), the VLDB 10 Year Best Paper Award (1999) and numerous IBM awards, Mohan was elected to the US and Indian National Academies of Engineering (2009) and named an IBM Master Inventor (1997). This Distinguished Alumnus of IIT Madras (1977) received his PhD at the University of Texas at Austin (1981). He is an inventor of 50 patents. He is currently focused on Blockchain, Big Data and HTAP technologies ( <http://bit.ly/CMbcDB> , <http://bit.ly/CMgMDS> ). For over 2 years, he has been an evangelist for permissioned blockchains and the myth buster of permissionless blockchains. Since 2016, Mohan has been a Distinguished Visiting Professor of China's prestigious Tsinghua University. He has served on the advisory board of IEEE Spectrum, and on numerous conference and journal boards. Mohan is a frequent speaker in North America, Europe and Asia, and has given talks in 40 countries. He is very active on social media and has a huge network of followers. More information can be found in the Wikipedia page at <http://bit.ly/CMwIkP>

**14:30 - 15:00**

**Session 3 (30 min each)**

**Benchmarking Databases "On-The-Go"**

*Authors: Carl Nuessle, Oliver Kennedy and Lukasz Ziarek*

**End-to-End Benchmarking of Deep Learning Platforms**

*Authors: Vincent Deuschle, Alexander Alexandrov and Volker Markl*

**15:30-16:00**

**Coffee Break**

**16:00 - 17:00**

**Panel: Role of the TPC in the Cloud Age**

**17:00 - 18:00**

**Session 4 (30 min each)**

**Benchmarking Database Cloud Services**

*Authors: Manfred Drozd*

**Use Machine Learning Methods to Predict Performance of Storage Devices**

*Authors: Yingxuan Zhu*

**18:00-18:30**

**Closing Remarks by Meikel Poess**

**FRIDAY – August 30 (Santa Anita)****BLOCKCHAIN AND DISTRIBUTED LEDGER (BCDL)****Chairs:** C. Mohan (IBM) and Beng Chin Ooi (National University of Singapore)*Please see workshop website for the most up to date schedule <https://bcdl.comp.nus.edu.sg/>***9:00 - 9:10****Opening (C. Mohan, IBM)****9:10 - 10:30****Keynote: Christopher Ferris, IBM.****Permissioned Blockchain is the Future - Public or Private**

**Abstract:** There's a raging debate in the blockchain community about whether permissioned blockchain is a true blockchain. Some would argue that only Bitcoin is a true blockchain (watch this SXSW Deathmatch 2019 <http://bit.ly/2GxXSeQ>). Others argue that public blockchain (implying permissionless) is the future, and that private blockchain (implying permissioned) is destined to become anachronistic. There are many factors at play. Many uses of blockchain require privacy of data beyond that which mere encryption can provide. Most financial applications of blockchain will need to conform with regulatory regimens such as KYC and AML that require that participants to transactions be identifiable. Then, there is the matter of performance at scale and finality which are challenges at best with most permissionless (public) blockchain platforms. In this talk, Christopher Ferris, IBM Fellow and CTO Open Technology, will present an argument that the question is not "public or private blockchain" but rather that blockchain solutions embraced by governments and enterprises will, of necessity, be permissioned.

**Bio:** An IBM Fellow, Chris is recognized as an expert in the architecture, design and engineering of distributed systems, and in open standards and open source development. Chris has been involved in the architecture, design, and engineering of distributed systems for most of his 39+ year career in IT. He has been actively engaged in open standards and open source development since 1999. Chris was one of the earliest to embrace Java to implement a three-tiered application architecture and was deeply involved with Microsoft in shaping the XML and Web services standards, authoring or editing a number of the WS-\* standards. He has overall responsibility for setting the strategy and overseeing the execution of IBM's strategic open source and standards initiatives. He represents IBM on the Hyperledger Technical Steering Committee, is a maintainer of the Hyperledger Fabric project.

**10:30 - 11:00****Coffee Break****11:00 - 12:30****Session 1****Consensus in Enterprise and Financial Blockchains: Assumptions and Challenges.***Authors: Hank Korth, LeHigh University.*

**Towards Global Asset Management in Blockchain Systems**

*Authors: Victor Zakhary, UCSB; Mohammad Javad Amiri, UCSB; Sujaya Maiyya, UCSB; Divy Agrawal, UCSB; Amr El Abbadi, UCSB.*

**Blockchain-based access audit for cross-site genomic data sharing**

*Authors: Li Xiong (Emory University)*

**12:30 - 14:00**

**Lunch Break**

**14:00 - 15:30**

**Session 2**

**Trusted Wide-Area and Edge Systems Using Hierarchical Permissioned Blockchain.**

*Authors: Faisal Nawab, UCSC; Mohammad Sadoghi, UCSC; Abhishek Singh, UCSC.*

**Reasoning about the Future in Blockchain Databases.**

*Authors: Sara Cohen, The Hebrew University of Jerusalem; Adam Rosenthal, The Hebrew University; Aviv Zohar, The Hebrew University.*

**Understanding the Scalability of Hyperledger Fabric.**

*Authors: Nguyen Minh Quang, NUS; Dumitrel Loghin, NUS; Tien Tuan Anh Dinh, SUTD.*

**15:30 - 16:00**

**Coffee Break**

**16:00 - 17:30**

**Panel: Trends and Directions in Blockchain Technology**

*Moderator: C. Mohan, IBM*

**Participants:** Mic Bowman, Intel; Bhaskar Krishnamachari, USC; Gabriela Ruberg, Central Bank of Brazil; Li Xiong, Emory University.



**FRIDAY – August 30 (Emerald Bay + Santa Monica)****BIG DATA OPEN SOURCE SYSTEMS (BOSS)**

**Chairs:** Pinar Tözün (IT University of Copenhagen) and Emanuel Zgraggen (MIT)

*Please see workshop website for the most up to date schedule*  
<http://boss.dima.tu-berlin.de/>

**9:00 - 9:20****Intro, Flash Talks (3 mins for each parallel tutorial)****9:20 - 10:30****Tutorial: Tensor Flow Extended (TFX): A platform for end-to-end ML in production (Neoklis Polyzotis (Google, USA))****10:30 - 11:00****Coffee Break****11:00 - 13:00****Parallel Tutorials I (1.5-2h based on tutorial, all but TFX)****13:00 - 14:00****Lunch Break****14:00 - 15:30****Parallel Tutorials II (1.5-2h based on tutorial, repetitions of the morning tutorials)****15:30 - 16:00****Coffee Break****16:00 - 16:30****Parallel Tutorials II Continued**

\*\*\*\*\*Parallel Tutorials \*\*\*\*\*

**Cloudberry for Big Data Visualization**

*Authors: Sadeem Alsudaïs, Qiushi Bai, and Chen Li (UC Irvine, USA)*

**Building massive virtual data warehouses with Apache Drill**

*Authors: Patrick Holl (TU Munich, Germany)*

**Storage Tuning and Programming for Data-Intensive Systems with OX and Open-Channel SSDs**

*Authors: Philippe Bonnet, Ivan Luiz Picoli (IT University of Copenhagen, Denmark)*

**From Zero to Hero with Apache Kudu**

*Authors: Andrew Wong (Cloudera, USA)*

**Apache Arrow**

*Authors: Wes McKinney (Ursa Labs & RStudio, USA)*

**FRIDAY – August 30 (Mt. Washington)****CONVERSATIONAL ACCESS TO DATA (CAST)**

**Chairs:** H.V Jagdish (University of Michigan), Georgia Koutrika (Athena Research Center), and Fatma Özcan (IBM Research - Almaden)

Please see workshop website for the most up to date schedule <https://cast19.athenarc.gr/>

**8:45 - 9:00****Welcome and Introduction**

**Keynote:** *Nikola Mrkšić (Co-founder and CEO, PolyAI)*

**9:00 - 10:00****Deploying a Conversational AI platform for Customer Support**

**Abstract:** PolyAI is a London-based startup with a leading machine learning platform for conversational agents. The deployed AI agents understand users, hold conversations without getting confused and can easily scale to new use cases or other languages. This talk will present the machine learning techniques that underpin the PolyAI platform and results from its early deployments in contact centre environments.

**Bio:** Nikola Mrkšić is the CEO and Co-Founder of PolyAI, a London-based Conversational AI company. Before starting PolyAI, Nikola worked with the Apple Siri team in Cambridge, and he was the first engineer at VocalIQ, a dialogue systems startup acquired by Apple. He did a PhD at Cambridge, working with Professor Steve Young at the Dialogue Systems Group.

**10:00 - 10:30****Session 1****Answering Complex Queries with Heterogeneous Structured Knowledge Sources extracted from Text**

*Authors: Nikita Bhutani (University of Michigan, Ann Arbor, Michigan)*

**10:30 - 11:00****Coffee Break****11:00 - 12:30****Session 2****Leveraging Human Learning in Interactive Data Exploration**

*Authors: Sanad Saha (Oregon State University), Arash Termehchy (Oregon State University), and Leilani Battle (University of Maryland)*

**DBPal: Weak Supervision for Learning a Natural Language Interface to Databases**

*Authors: Nathaniel Weir (Brown University), Andrew Crotty (Brown University), Alex Galakatos (Brown University), Amir Ilkechi (Brown University), Shekar Ramaswamy (Brown University), Rohin Bhushan (Brown University), Ugur Cetintemel (Brown University), P. Ajie Utama (TU Darmstadt), Nadja Geisler (TU Darmstadt), Benjamin Hättasch (TU Darmstadt), Seffen Eger (TU Darmstadt), and Carsten Binnig (TU Darmstadt)*

**Disambiguating Natural Language Queries with Tuples**

*Authors: Christopher Baik (University of Michigan); Zhongjun Jin (University of Michigan); Michael Cafarella (University of Michigan)*

**12:30 - 14:00****Lunch Break****14:00 - 15:00**

**Keynote: Dilek Hakkani-Tur (Senior Principal Scientist at Amazon Alexa AI)**

**Conversational Machines: Bridging the chasm between task-oriented and social conversations**

**Abstract:** Conversational systems generally fall into two categories: task-oriented and social bots. Task-oriented systems aim to help users accomplish a specific task through multi-turn interactions, whereas socialbots focus on engaging and natural open-domain conversations. In natural interactions, even when conversation participants have a task or goal in mind, they can say things that are out of the boundaries of that task domain, and similarly they can switch to a task during a chitchat conversation. Hence, the ability to engage in knowledgeable social interactions and gracefully transition back and forth to the task is crucial for enabling natural conversations.

In this talk, I'll summarize our recent work in both fronts, focusing on the convergence of approaches for the two categories of conversational systems. Starting with task-oriented interactions, I'll present our approach for bootstrapping task-oriented dialogue systems from simulated seeker-provider conversations and dialogue state tracking using generate-and-copy mechanisms. This will be followed by a summary of learnings from previous Alexa Prize challenges and progression of our work as we approach the next challenge.

**Bio:** Dilek Hakkani-Tür is a senior principal scientist at Amazon Alexa AI focusing on enabling natural dialogues with machines. Prior to joining Amazon, she was leading the dialogue research group at Google, and worked as a principal researcher at Microsoft Research, International Computer Science Institute and AT&T Labs-Research. She received her BSc degree from Middle East Technical Univ. and MSc and PhD degrees in Computer Science from Bilkent University. Her research interests include conversational AI, natural language and speech processing, spoken dialogue systems, and machine learning for language processing. She has over 70 patents that were granted and co-authored more than 200 papers in natural language and speech processing. She is the Editor-in-Chief of the IEEE/ACM Transactions on Audio, Speech and Language Processing, and a fellow of the IEEE and ISCA.

**15:00 - 15:30****Session 3****Building a Hotel Concierge Bot: an industrial case study**

*Authors: Behzad Golshan (Megagon Labs), George Mihaila (Megagon Labs), Chen Chen (Megagon Labs), Jonathan Engel (Megagon Labs), Alon Halevy (Megagon Labs), Yoshihiko Suhara (Megagon Labs), Wang-Chiew Tan (Megagon Labs), Michael Matuschek (TrustYou)*

**15:30 - 16:00****Coffee Break****16:00 - 17:30**

**Panel: Industry panel on "Conversational Access to Data and Enterprise Chatbots"**

**17:30****Closing Remarks**

**FRIDAY – August 30 (San Jose)****DATA MANAGEMENT AND ANALYTICS FOR MEDICINE AND HEALTHCARE (DMAH)**

**Chairs:** Fusheng Wang (Stony Brook University), Gang Luo (University of Washington), Yanhui Liang (Google Inc), and Alevtina Dubovitskaya (Lucerne University of Applied Sciences and Arts and Swisscom)

Please see workshop website for the most up to date schedule <https://sites.google.com/site/vldb dmah2019/>

**9:00 - 10:30****Session 1: Database Enabled Biomedical Research****Patient Centric Data Integration for Improved Diagnosis and Risk Prediction**

*Authors: Hanie Samimi, Anne Hee Hiong Ngu and Jelena Tesic*

**An Architecture to Support Real-World Studies that Investigate the Autonomic Nervous System**

*Authors: Danielle Groat, Ramkiran Gouripeddi, Yu Kuei Lin and Julio Facelli*

**Comparison of Approaches for Querying Chemical Compounds**

*Authors: Vojtech Sipek and Irena Holubova*

**10:30 - 11:00****Coffee Break****11:00-12:30****Session 2: AI for Healthcare****Differential Diagnosis of Heart Disease Using Decision Tree and Medical Knowledge in Emergency Department**

*Authors: Diyang Xue, Adam Frisch, and Daqing He*

**Deep Autoencoder Based Neural Networks for Coronary Heart Disease Risk Prediction**

*Authors: Tsatsral Amarbayasgalan, Keun Ho Ryu, Jong Yun Lee and Kwang Rok Kim*

**Towards Automated Hypothesis Testing in Neuroscience**

*Authors: Daniel Garijo, Shobeir Fakhraei, Varun Ratnakar, Qifan Yang, Hanna Endrias, Yibo Ma, Regina Wang, Michael Bornstein, Joanna Bright, Yolanda Gil and Neda Jahanshad*

**12:30-13:30****Lunch Break****14:00-15:30****Session 3: Knowledge Discovery from Unstructured Biomedical Data**

Keynote: *Prof. Wei Wang*  
**Transforming unstructured biomedical text to structured knowledge**

**Bio:** Dr. Wei Wang is the Leonard Kleinrock Chair Professor in Computer Science and Computational Medicine at University of California, Los Angeles

**Training Set Expansion Using Word Embeddings for Korean Medical Information Extraction**

*Authors: Young-Min Kim*

**15:30-16:00**

**Coffee Break**

**16:00-17:30**

**Session 4: Blockchain and Privacy-preserving Data Management**

**Intelligent Healthcare Data Management using Blockchain: Current Limitation and Future Research Agenda**

*Authors: Alevtina Dubovitskaya, Petr Novotny, Scott Thiebess, Ali Sunyae, Michael Ignaz Schumacher, Zhigang Xu and Fusheng Wang*

Keynote: *Prof. Lucila Ohno-Machado*  
**Sharing data while protecting privacy**

**Bio:** Dr. Lucila Ohno-Machado is Associate Dean for Informatics and Technology, and the Founding Chair of the Health System Department of Biomedical Informatics at University of California, San Diego

**FRIDAY – August 30 (San Diego)****POLYSTORES INCLUDING MANAGING PRIVACY, SECURITY AND POLICY FOR HETEROGENOUS DATA (POLY)**

**Chairs:** Vijay Gadepally (MIT), Tim Mattson (Intel), Mike Stonebraker (MIT)

Please see workshop website for the most up to date schedule  
<https://sites.google.com/view/poly19/>

**8:45 - 9:00****Welcome and Overview of Day by Workshop Chairs****9:00 - 10:40****Session 1: Privacy, Security and/or Policy Issues for Heterogenous Data****Invited Talk 1: Systems Challenges of Taking Privacy Seriously**

*Authors: Daniel J. Weitzner (Founding Director, MIT Internet Policy Research Initiative)*

**WIP: Data Capsule: A New Paradigm for Automatic Compliance of Data Privacy Regulations**

*Authors: Lun Wang (University of California, Berkeley); Joseph Near (University of Vermont); Neel Somani (University of California, Berkeley); Peng Gao (University of California, Berkeley); Andrew Low (University of California, Berkeley); David Dao (ETH Zurich); Dawn Song (UC Berkeley)*

**DATUMDB: A Data Protection Database Proposal**

*Authors: Tim Kraska (MIT); Michael Stonebraker (MIT); Michael Brodie (MIT); Sacha Servan-Schreiber (MIT); Daniel J. Weitzner (MIT)*

**Position: GDPR Compliance by Construction**

*Authors: Malte Schwarzkopf (MIT CSAIL); Eddie Kohler (Harvard University); Frans Kaashoek (MIT CSAIL); Robert Morris (MIT CSAIL)*

**10:40 - 11:00****Coffee Break****11:00 - 12:40****Session 1 (continuation)****Invited Talk 2: Google, AI, Devices, Services, and Privacy**

*Authors: Blaise Aguera y Arcas (Google Inc.)*

**Understanding Information Flow in Distributed Systems with Data Provenance**

*Authors: Thomas F. J.-M. Pasquier (University of Bristol); David Ebers (University of Otago); Margo Seltzer (University of British Columbia)*

**Position: Privacy and Policy in Polystores: A Data Management Research Agenda**

*Authors: Joshua Kroll (UC Berkeley); Nitin Kohli (UC Berkeley School of Information); Paul Laskowski (UC Berkeley School of Information)*

**Regular: Analyzing GDPR Compliance Through the Lens of Privacy Policy**

*Authors: Jayashree Mohan (UT Austin); Melissa Wasserman (School of Law, UT Austin); Vijay Chidambaram (UT Austin and VMWare)*

**12:40 - 13:30**

**Lunch Break**

**13:30 - 17:00**

**Session 2: Building Polystore Systems**

**Invited Talk 3: Unifying the Querying of Structured and Semistructured data with PartiQL**

*Authors: Yannis Papakonstantinou (UCSD/Amazon)*

**Learning How to Optimize Data Access in Polystores**

*Authors: Antonio Maccioni (Collective[i]), Riccardo Torlone (Roma Tre University)*

**Midas: Towards an interactive data catalog**

*Authors: Patrick Holl (Technical University of Munich); Kevin Gofßling (TU Munich)*

**Evolution Management of Multi- Model Data**

*Authors: Irena Holubova (Charles University); Meike Klettke (Universität Rostock); Uta Störl (University of Applied Sciences Darmstadt)*

**SKOD: A Framework for Situational Knowledge on Demand**

*Authors: Servio Palacios (Purdue University); K M A Solaiman (Purdue University); Pelin Angin (METU, Turkey); Alina V Nesen (Purdue University); Bharat Bhargava (Purdue University); Zachary Collins (MIT); Aaron Sipser (MIT); Michael Stonebraker (MIT); Jim MacDonald (NGC)*

**15:30 - 16:00**

**Coffee Break**

**16:00 - 17:00**

**Session 2 (continuation)**

**Invited Talk 4: The Journey Towards Polystore Systems at Facebook**

*Authors: Pedro Pedreira, Facebook Inc.*

**Development of a Polystore Data Management System for an Evolving Big Scientific Data Archive**

*Authors: Manoj Poudel (The University of Aizu); Rashmi P Sarode (The University of Aizu); Subhash Bhalla (IIT Delhi)*

**17:00 - 17:30**

**Closing Remarks**

**FRIDAY – August 30 (Avalon)****VERY LARGE INTERNET OF THINGS (VLIOT)**

**Chairs:** Sven Groppe (University of Lübeck, Germany) and Markus Endler (PUC Rio, Brasil)

Please see workshop website for the most up to date schedule

<https://www.ifis.uni-luebeck.de/~groppe/vliot/2019>

**9:00 - 10:30**

**Session 1 (Welcome, Keynote/Cloud-IoT Systems)**

**Short Welcome :** Sven Groppe and Markus Endler

Keynote: *Sumi Helal (Lancaster University)*

**Energy Savings in Very Large Cloud-IoT Systems**

**Abstract:** Opposite to the original cloudlet approach in which an edge is utilized to bring the cloud and its benefits closer to the applications, in cloud- and edge-connected IoT systems where the applications are deployed and run in the cloud, we exploit the edge somewhat differently, either by bringing the physical world and its data up closer to the cloud or by caching parts of the applications down closer to the physical world. Aggressive optimizations seeking substantial IoT energy savings are needed to maintain the scalability of large-scale IoT deployments and to stay within cloud cost constraints (avoiding costly elasticity when working with a budget limit). In this keynote, we present a novel optimization approach that relies on the simple principle of minimizing all movements: movements of data from the IoT up to the Edge and Cloud, and movements of application fragments from the cloud down to the edge and the IoT itself. Our approach is novel in that it involves and utilizes the dynamic characteristics and variability of both the data and applications simultaneously. Another novelty of our approach is the definition and use of ‘sentience-efficiency’ as a precursor to ‘energy-efficiency’ for achieving truly aggressive savings in energy. We present our bi-directional optimization approach and its implementation in terms of algorithms within an architecture we name the cloud-edge-beneath architecture (CEB). We present a performance evaluation study to measure the impact of our optimization approach on energy saving.

**Bio:** Abdelsalam (Sumi) Helal received the Ph.D. degree in computer sciences from Purdue University, West Lafayette, IN, USA. He is currently professor and the Chair in Digital Health, School of Computing and Communications, and the Division of Health Research, Lancaster University, UK. Before joining Lancaster University, he was professor in the department of Computer and Information Science and Engineering, University of Florida, USA, where he directed the Mobile and Pervasive Computing Laboratory and the Gator Tech Smart House. His research interests span pervasive systems, the Internet of Things, smart spaces, with applications to digital health and assistive technologies for successful aging and independence.

**Data-Centric Resource Management in Edge-Cloud Systems for the IoT**

*Authors: Igor Leão dos Santos, Flávia C. Delicato, Paulo F. Pires, Marcelo Pitanga Alves, Ana Oliveira, Tiago Salviano Calmon*

**10:30 - 11:00**

**Coffee Break**



**11:00 - 12:30****Session 2****Online Replication Strategies for Distributed Data Stores***Authors: Niklas Semmler, Georgios Smaragdakis, Anja Feldmann***Understanding the Performance of Software Defined Wireless Sensor Networks Under Denial of Service Attack***Authors: Gustavo A. Nunez Segura, Cintia Borges Margi, Arsenia Chorti***IoT Data Imputation with Incremental Multiple Linear Regression***Authors: Tao Peng, Sana Sellami, Omar Boucelma***12:30 - 14:00****Lunch Break****14:00 - 15:30****Session 3 (Panel and Visions)**

**Panel:** IoT as core technology for SmartEverything, Data Science and Machine Learning

**Participants:** Flávia C. Delicato (Universidade Federal do Rio de Janeiro), Cintia Borges Margi (Universidade de São Paulo), Sumi Helal (Lancaster University), Satyajant Misra (New Mexico State University)

**Abstract:** If the IoT infrastructure is the nervous system of a cyber-physical system, Data Science is the knowledge construction, and Machine Learning is the brain, how can we be sure that we are collecting and processing all bits of information to build really smart, adaptive and human-friendly systems?

**Towards a Large Scale IoT through Partnership, Incentive, and Services: A Vision, Architecture, and Future Directions***Authors: Gowri Sankar Ramachandran, Bhaskar Krishnamachari***Distributed Data-Gathering and -Processing in Smart Cities: An Information-Centric Approach***Authors: Reza Tourani, Abderrahmen Mtibaa, Satyajant Misra***15:30 - 16:00****Coffee Break****16:00 - 17:40****Session 4****Leveraging Application Development for the Internet of Mobile Things***Authors: Felipe Oliveira Carvalho, Markus Endler, Francisco Silva e Silva***Integrating a smart city testbed into a large scale heterogeneous federation of Future Internet experimentation facilities: the SmartSantander approach***Authors: Pablo Sotres, Jorge Lanza, Juan Ramón Santana, and Luis Sanchez***Experimentation and Analysis of Ensemble Deep Learning in IoT Applications***Authors: Taylor Mauldin, Anne H. Ngu, Vangelis Metsis, Marc E. Canby, Jelena Tesic***Data lifetime estimation in a multicast-based CoAP proxy***Authors: Jelena Mistic, Vojislav Mistic, Xiaolin Chang*

# DEMO SESSIONS

The number assigned to each demo identifies its table and easel in the Hollywood ballroom. For example, Demo 5 in Group B is assigned to Table and Easel numbered 5.

## Demo Session: Group A

### A.1 - Synergistic Graph and SQL Analytics Inside IBM Db2

Authors: Yuanyuan Tian (IBM Research), Sui Jun Tong (IBM Research), Mir Hamid Pirahesh (IBM Research), Wen Sun (IBM Research), En Liang Xu (IBM Research), and Wei Zhao (IBM Research)

### A.2 - ITAA: An Intelligent Trajectory-driven Outdoor Advertising Deployment Assistant

Authors: Yipeng Zhang (RMIT University), Zhifeng Bao (RMIT University), Songsong Mo (Wuhan University), Yuchen Li (SMU), and Yanghao Zhou (RMIT University)

### A.3 - Buckle: Evaluating Fact Checking Algorithms Built on Knowledge Bases

Authors: Viet-Phi Huynh (Eurecom) and Paolo Papotti (Eurecom)

### A.4 - A Query System for Efficiently Investigating Complex Attack Behaviors for Enterprise Security

Authors: Peng Gao (UC Berkeley), Xusheng Xiao (Case Western Reserve University), Zhichun Li (NEC Labs America), Kangkook Jee (UT Dallas), Fengyuan Xu (Nanjing University), Sanjeev R. Kulkarni (Princeton University), and Prateek Mittal (Princeton University)

### A.5 - ProgressiveDB – Progressive Data Analytics as a Middleware

Authors: Lukas Berg (TU Darmstadt), Tobias Ziegler (TU Darmstadt), Carsten Binnig (TU Darmstadt), and Uwe Röhm (University of Sydney)

### A.6 - doppioDB 2.0: Hardware Techniques for Improved Integration of Machine Learning into Databases

Authors: Kaan Kara (ETH Zurich), Zeke Wang (ETH Zurich), Ce Zhang (ETH Zurich), and Gustavo Alonso (ETH Zurich)

### A.7 - NashDB: Fragmentation, Replication, and Provisioning using Economic Methods

Authors: Ryan Marcus (Brandeis University), Chi Zhang (Brandeis University), Shuai Yu (Brandeis University), Geoffrey Kao (Brandeis University), and Olga Papaemmanouil (Brandeis University)

### A.8 - In-database Distributed Machine Learning: Demonstration using Teradata SQL Engine

Authors: Sandeep Singh Sandha (UCLA), Wellington Cabrera (Teradata Labs), Mohammed Al-Kateb (Teradata Labs), Sanjay Nair (Teradata Labs), and Mani Srivastava (UCLA)

### A.9 - SHOAL: Large-scale Hierarchical Taxonomy via Graph-based Query Coalition in E-commerce

*Authors: Zhao Li (Alibaba Group), Xia Chen (Alibaba Group, Southwest University), Xuming Pan (Alibaba Group), Pengcheng Zou (Alibaba Group), Yuchen Li (Singapore Management University), and Guoxian Yu (Southwest University)*

#### **A.10 - PriSTE: Protecting Spatiotemporal Event Privacy in Continuous Location-Based Services**

*Authors: Yang Cao (Kyoto University), Yonghui Xiao (Google Inc.), Li Xiong (Emory University), Liquan Bai (Emory University), and Masatoshi Yoshikawa (Kyoto University)*

#### **A.11 - HERMIT in Action: Succinct Secondary Indexing Mechanism via Correlation Exploration**

*Authors: Yingjun Wu (IBM Research - Almaden), Jia Yu (Arizona State University), Yuanyuan Tian (IBM Research - Almaden), Richard Sidle (IBM), and Ronald Barber (IBM Research - Almaden)*

#### **A.12 - DISPERS: Securing Highly Distributed Queries on Personal Data Management Systems**

*Authors: Julien Loudet (Cozy Cloud, INRIA Saclay, University of Versailles), Iulian Sandu-Popa (University of Versailles, INRIA Saclay), and Luc Bouganim (INRIA Saclay, University of Versailles)*

#### **A.13 - Demonstration of Krypton: Optimized CNN Inference for Occlusion-based Deep CNN Explanations**

*Authors: Allen Ordookhanians (University of California, San Diego), Xin Li (University of California, San Diego), Supun Nakandala (University of California, San Diego), and Arun Kumar (University of California, San Diego)*

#### **A.14 - Spade: A Modular Framework for Analytical Exploration of RDF Graphs**

*Authors: Yanlei Diao (LIX (UMR 7161, CNRS and Ecole polytechnique), University of Massachusetts Amherst), Pawel Guzewicz (LIX (UMR 7161, CNRS and Ecole polytechnique), INRIA), Ioana Manolescu (LIX (UMR 7161, CNRS and Ecole polytechnique), INRIA), and Mirjana Mazuran (LIX (UMR 7161, CNRS and Ecole polytechnique), INRIA)*

#### **A.15 - Making an RDBMS Data Scientist Friendly: Advanced In-database Interactive Analytics with Visualization Support**

*Authors: Joseph Vinish D'silva (McGill University), Florestan De Moor (McGill University), and Bettina Kemme (McGill University)*

#### **A.16 - UDAO: A Next-Generation Unified Data Analytics Optimizer**

*Authors: Khaled Zaouk (Ecole Polytechnique), Fei Song (Ecole Polytechnique), Chenghao Lyu (University of Massachusetts Amherst), Arnab Sinha (Ecole Polytechnique), Yanlei Diao (Ecole Polytechnique, University of Massachusetts Amherst), and Prashant Shenoy (University of Massachusetts Amherst)*

#### **A.17 - Dietcoin: Hardening Bitcoin Transaction Verification Process For Mobile Devices**

*Authors: Davide Frey (Univ Rennes, Inria, CNRS, IRISA), Marc X. Makkes (Vrije Universiteit), Pierre-Louis Roman (Univ Rennes, Inria, CNRS, IRISA), François Taïani (Univ Rennes, CNRS, Inria, IRISA), and Spyros Voulgaris (Athens University of Economics and Business)*

#### **A.18 - COVIZ: A System for Visual Formation and Exploration of Patient Cohorts**

*Authors: Cicero A. L. Pahins (UFRGS), Behrooz Omidvar-Tehrani (Univ. Grenoble Alpes, CNRS, LIG), Sihem Amer-Yahia (Univ. Grenoble Alpes, CNRS, LIG), Valérie Siroux (Univ. Grenoble Alpes, CNRS, LIG, Institute for Advanced Biosciences and Inserm), Jean-Louis Pepin (Univ. Grenoble Alpes, CNRS, LIG, Grenoble University Hospital), Jean-Christian Borel (AGIR-à-Dom), and João Comba (UFRGS)*

#### **A.19 - WiClean: A System for Fixing Wikipedia Interlinks Using Revision History Patterns**

*Authors: Stephan Goldberg (Tel Aviv University), Tova Milo (Tel Aviv University), Slava Novgorodov (eBay Research), and Kathy Razmadze (Tel Aviv University)*

#### **A.20 - Datalignment: Ontology Schema Alignment Through Datalog Containment**

*Authors: Daniel Deutch (Tel Aviv University), Evgeny Marants (Tel Aviv University), and Yuval Moskovitch (Tel Aviv University)*

#### **A.21 - FishStore: Fast Ingestion and Indexing of Raw Data**

*Authors: Badrish Chandramouli (Microsoft Research), Dong Xie (University of Utah), Yinan Li (Microsoft Research), and Donald Kossmann (Microsoft Research)*

#### **A.22 - Raptor: Large Scale Analysis of Big Raster and Vector Data**

*Authors: Samridhhi Singla (UC Riverside), Ahmed Eldawy (UC Riverside), Rami Alghamdi (University of Minnesota, Twin Cities), and Mohamed F. Mokbel (Qatar Computing Research Institute)*

#### **A.23 - Tplex: Robust, Efficient Analytics When Python Rules**

*Authors: Leonhard F. Spiegelberg (Brown University) and Tim Kraska (Massachusetts Institute of Technology)*

#### **A.24 - Ease.ml/ci and Ease.ml/meter in Action: Towards Data Management for Statistical Generalization**

*Authors: Cedric Renggli (ETH Zurich), Frances Ann Hubis (ETH Zurich), Bojan Karlaš (ETH Zurich), Kevin Schawinski (Modulos AG), Wentao Wu (Microsoft Research, Redmond), and Ce Zhang (ETH Zurich)*

### **Demo Session: Group B**

#### **B.1 - BlackMagic: Automatic Inlining of Scalar UDFs into SQL Queries with Froid**

*Authors: Karthik Ramachandra (Microsoft Research India) and Kwanghyun Park (Microsoft Gray Systems Lab)*

#### **B.2 - Data Civilizer 2.0: A Holistic Framework for Data Preparation and Analytics**

*Authors: El Kindi Rezig (MIT CSAIL), Lei Cao (MIT CSAIL), Michael Stonebraker (MIT CSAIL), Giovanni Simonini (MIT CSAIL), Wenbo Tao (MIT CSAIL), Samuel Madden (MIT CSAIL), Mourad Ouzzani (Qatar Computing Research Institute, HBKU), Nan Tang (Qatar Computing Research Institute, HBKU), and Ahmed K. Elmagarmid (Qatar Computing Research Institute, HBKU)*

#### **B.3 - VISE: Vehicle Image Search Engine with Traffic Camera**

*Authors: Hyewon Choi (University of Toronto), Erkang Zhu (University of Toronto), Arsalan Bangash (University of Toronto), and Renée J. Miller (Northeastern University)*

#### **B.4 - Stateful Functions as a Service in Action**

*Authors: Adil Akhter (ING), Marios Fragkoulis (Delft University of Technology), and Asterios Katsifodimos (Delft University of Technology)*

#### **B.5 - GALO: Guided Automated Learning for re-Optimization**

*Authors: Guilherme Damasio (Ontario Tech University, IBM Centre for Advanced Studies), Spencer Bryson (Ontario Tech University, IBM Centre for Advanced Studies), Vincent Corvinelli (IBM Ltd), Parke Godfrey (York University, IBM Centre for Advanced Studies), Piotr Mierzejewski (IBM Ltd), Jaroslaw Szlichta (Ontario Tech University, IBM Centre for Advanced Studies), and Calisto Zuzarte (IBM Ltd)*

#### **B.6 - CAPE: Explaining Outliers by Counterbalancing**

*Authors: Zhengjie Miao (Duke University), Qitian Zeng (IIT), Chenjie Li (IIT), Boris Glavic (IIT), Oliver Kennedy (SUNY Buffalo), and Sudeepa Roy (Duke University)*

**B.7 - Cleanits: A Data Cleaning System for Industrial Time Series**

*Authors: Xiaou Ding (Harbin Institute of Technology), Hongzhi Wang (Harbin Institute of Technology), Jiakuan Su (Harbin Institute of Technology), Zijue Li (Harbin Institute of Technology), Jianzhong Li (Harbin Institute of Technology), and Hong Gao (Harbin Institute of Technology)*

**B.8 - SystemER: A Human-in-the-loop System for Explainable Entity Resolution**

*Authors: Kun Qian (IBM Research - Almaden), Lucian Popa (IBM Research - Almaden), and Prithviraj Sen (IBM Research - Almaden)*

**B.9 - PRIMAT: A Toolbox for Fast Privacy-preserving Matching**

*Authors: Martin Franke (Leipzig University & ScaDS Dresden/Leipzig), Ziad Sehili (Leipzig University & ScaDS Dresden/Leipzig), and Erhard Rahm (Leipzig University & ScaDS Dresden/Leipzig)*

**B.10 - Flash in Action: Scalable Spatial Data Analysis Using Markov Logic Networks**

*Authors: Ibrahim Sabek (University of Minnesota), Mashaal Musleh (University of Minnesota), and Mohamed Mokbel (Qatar Comp. Research Inst., HBKU)*

**B.11 - I Can't Believe It's Not (Only) Software! Bionic Distributed Storage for Parquet Files**

*Authors: Lucas Kuhring (IMDEA Software Institute) and Zsolt István (IMDEA Software Institute)*

**B.12 - SparkCruise: Handsfree Computation Reuse in Spark**

*Authors: Abhishek Roy (Microsoft), Alekh Jindal (Microsoft), Hiren Patel (Microsoft), Ashit Gosalia (Microsoft), Subru Krishnan (Microsoft), and Carlo Curino (Microsoft)*

**B.13 - DPSAaS: Multi-Dimensional Data Sharing and Analytics as Services under Local Differential Privacy**

*Authors: Min Xu (University of Chicago), Tianhao Wang (Purdue University), Bolin Ding (Alibaba Group), Jingren Zhou (Alibaba Group), Cheng Hong (Alibaba Group), and Zhicong Huang (Alibaba Group)*

**B.14 - IHCS: An Integrated Hybrid Cleaning System**

*Authors: Congcong Ge (Zhejiang University), Yunjun Gao (Zhejiang University, Alibaba-Zhejiang University Joint Institute of Frontier Technologies), Xiaoye Miao (Zhejiang University), Lu Chen (Aalborg University), Christian S. Jensen (Aalborg University), and Ziyuan Zhu (Zhejiang University)*

**B.15 - CAPRIO: Graph-based Integration of Indoor and Outdoor Data for Path Discovery**

*Authors: Constantinos Costa (University of Pittsburgh), Xiaoyu Ge (University of Pittsburgh), and Panos K. Chrysanthis (University of Pittsburgh)*

**B.16 - LensXPlain: Visualizing and Explaining Contributing Subsets for Aggregate Query Answers**

*Authors: Zhengjie Miao (Duke University), Andrew Lee (Duke University), and Sudeepa Roy (Duke University)*

**B.17 - Juneau: Data Lake Management for Jupyter**

*Authors: Yi Zhang (University of Pennsylvania) and Zachary G. Ives (University of Pennsylvania)*

**B.18 - ApproxML: Efficient Approximate Ad-Hoc ML Models Through Materialization and Reuse**

*Authors: Sona Hasani (University of Texas at Arlington), Faezeh Ghaderi (University of Texas at Arlington), Shohedul Hasan (University of Texas at Arlington), Saravanan Thirumuruganathan (QCRI, HBKU), Abolfazl Asudeh (University of Illinois at Chicago), Nick Koudas (University of Toronto), and Gautam Das (University of Texas at Arlington)*

**B.19 - Flare & Lantern: Efficiently Swapping Horses Midstream**

*Authors: Grégory Essertel (Purdue University), Ruby Y. Tahboub (Purdue University), Fei Wang (Purdue University), James Decker (Purdue University), and Tiark Rompf (Purdue University)*

**B.20 - Trinity: An Extensible Synthesis Framework for Data Science**

*Authors: Ruben Martins (Carnegie Mellon University), Jia Chen (UT Austin), Yanju Chen (UCSB), Yu Feng (UCSB), and Isil Dillig (UT Austin)*

**B.21 - PSynDB: Accurate and Accessible Private Data Generation**

*Authors: Zhiqi Huang (Univ. of Massachusetts, Amherst), Ryan McKenna (Univ. of Massachusetts, Amherst), George Bissias (Univ. of Massachusetts, Amherst), Jerome Miklau (Univ. of Massachusetts, Amherst), Michael Hay (Colgate University), and Ashwin Machanavajjhala (Duke University)*

**B.22 - AggChecker: A Fact-Checking System for Text Summaries of Relational Data Sets**

*Authors: Saehan Jo (Cornell University), Immanuel Trummer (Cornell University), Weicheng Yu (Cornell University), Xuezi Wang (Google Research), Cong Yu (Google Research), Daniel Liu (Cornell University), and Niyati Mehta (Cornell University)*

**B.23 - GRANO: Interactive Graph-based Root Cause Analysis for Cloud-Native Distributed Data Platform**

*Authors: Hanzhang Wang (eBay Inc.), Phuong Nguyen (eBay Inc.), Jun Li (eBay Inc.), Selcuk Kopru (eBay Inc.), Gene Zhang (eBay Inc.), Sanjeev Katariya (eBay Inc.), and Sami Ben-Romdhane (eBay Inc.)*

**B.24 - PivotE: Revealing and Visualizing the Underlying Entity Structures for Exploration**

*Authors: Han Xueran (Renmin University of China), Jun Chen (Renmin University of China), Jiaheng Lu (University of Helsinki), Yueguo Chen (Renmin University of China), and Xiaoyong Du (Renmin University of China)*

# POSTERS OF VLDB JOURNAL PAPERS

*Posters of VLDB Journal Papers from 6/16/2018 to 6/15/2019.*

*The number assigned to each paper identifies its poster stand in the Jogging track. For example, VLDBJ:1 is assigned to stand labeled VLDBJ:1.*

## **VLDBJ:1 - Accelerating Pairwise SimRank Estimation over Static and Dynamic Graphs**

*Authors: Yue Wang (Hong Kong University of Science and Technology), Lei Chen (Hong Kong University of Science and Technology), Yulin Che (Hong Kong University of Science and Technology), and Qiong Luo (Hong Kong University of Science and Technology)*

## **VLDBJ:2 - Building Self-Clustering RDF Databases Using Tunable-LSH**

*Authors: Güneş Aluç (SAP SE), M. Tamer Özsu (University of Waterloo), and Khuzaima Daudjee (University of Waterloo)*

## **VLDBJ:3 - Scaling Distributed Transaction Processing and Recovery based on Dependency Logging**

*Authors: Chang Yao (National University of Singapore), Meihui Zhang (Beijing Institute of Technology), Qian Lin (National University of Singapore), Beng Chin Ooi (National University of Singapore), and Jiatao Xu (Tencent Inc.)*

## **VLDBJ:4 - A Framework for Efficient Multi-Attribute Movement Data Analysis**

*Authors: Fabio Valdés (FernUniversität in Hagen) and Ralf Hartmut Güting (FernUniversität in Hagen)*

## **VLDBJ:5 - Real-time Context-aware Social Media Recommendation**

*Authors: Xiangmin Zhou (RMIT University), Dong Qin (RMIT University), Lei Chen (Hong Kong University of Science and Technology), and Yanchun Zhang (Victoria University)*

## **VLDBJ:6 - A Survey of State Management in Big Data Processing Systems**

*Authors: Quoc-Cuong To (German Research Center for Artificial Intelligence), Juan Soto (DFKI & TU Berlin), and Volker Markl (DFKI & TU Berlin)*

## **VLDBJ:7 - Event Modeling and Mining: A Long Journey Towards Explainable Events**

*Authors: Xinhong Chen (City University of Hong Kong) and Qing Li (Hong Kong Polytechnic University)*

## **VLDBJ:8 - Comparing heuristics for graph edit distance computation**

*Authors: David B. Blumenthal (Free University of Bozen-Bolzano), Nicolas Boria (Normandie Université), Johann Gamper (Free University of Bozen-Bolzano), Sébastien Bougleux (Normandie Université), and Luc Brun (Normandie Université)*

## **VLDBJ:9 - VBTree - Forward Secure Conjunctive Queries over Encrypted Data for Cloud Computing**

*Authors: Zhiqiang Wu (Hunan University) and Kenli Li (Hunan University)*

**VLDBJ:10 - Cascade-Aware Partitioning of Large Graph Databases**

*Authors: Gunduz Vehbi Demirci (Bilkent Üniversitesi), Hakan Ferhatosmanoglu (University of Warwick), and Cevdet Aykanat (Bilkent Üniversitesi)*

**VLDBJ:11 - Summarizing Semantic Graphs: A Survey**

*Authors: Šejla Čebirić (Inria and LIX), François Goasdoué (Univ Rennes, Inria, CNRS, IRISA), Haridimos Kondylakis (FORTH-ICS), Dimitris Kotzinos (University of Paris-Seine, University of Cergy-Pontoise, ENSEA, CNRS), Ioana Manolescu (Inria and Lix), Georgia Troullinou (FORTH-ICS), and Mussab Zneika (University of Paris-Seine, University of Cergy-Pontoise, ENSEA, CNRS)*

**VLDBJ:12 - Location Prediction in Large-Scale Social Networks: An In-depth Benchmarking Study**

*Authors: Nur Al Hasan Haldar (The University of Western Australia), Jianxin Li (The University of Western Australia), Mark Reynolds (The University of Western Australia), Timos Sellis (Swinburne University of Technology), and Jeffrey Xu Yu (Chinese University of Hong Kong)*

**VLDBJ:13 - A Survey of Community Search Over Big Graphs**

*Authors: Yixiang Fang (Guangzhou University), Xin Huang (Hong Kong Baptist University), Lu Qin (University of Technology Sydney), Ying Zhang (University of Technology Sydney), Wenjie Zhang (University of New South Wales), Reynold Cheng (University of Hong Kong), and Xuemin Lin (University of New South Wales)*

**VLDBJ:14 - Microblogs Data Management: A Survey**

*Authors: Amr Magdy (University of California Riverside), Laila Abdelhafeez (University of California, Riverside), Yunfan Kang (University of California Riverside), Eric Ong (University of California Riverside), and Mohamed Mokbel (University of Minnesota)*

**VLDBJ:15 - Prescriptive Analytics: A Survey of Emerging Trends And Technologies**

*Authors: Davide Frazzetto (Aalborg Universitet), Thomas Dyhre Nielsen (Aalborg Universitet), Torben Bach Pedersen (Aalborg Universitet), and Laurynas Siksnys (Aalborg Universitet)*



# VLDB 2020



**August 31st - September 4th, 2020**

<https://vldb2020.org/>





**It's only  
impossible  
until it works**

**Ready to pioneer?  
Apply at [Amazon.jobs](https://amazon.jobs)**



# World's First "Self-Driving" Database



**Protects your Data from Cyber Attacks !**  
**Automatically Updates Security while Running**

**ORACLE®**

[oracle.com/selfdrivingdb](https://oracle.com/selfdrivingdb)

Copyright © 2017, Oracle and/or its affiliates. All rights reserved. Oracle and Java are registered trademarks of Oracle and/or its affiliates.

## TO MAKE IT EASY TO DO BUSINESS ANYWHERE

Alibaba Group enable businesses to transform the way they market, sell and operate and improve their efficiencies. We provide the technology infrastructure and marketing reach to help merchants, brands and other businesses to leverage the power of new technology to engage with their users and customers and operate in a more efficient way.

## Meet Alibaba Group@ VLDB 2019

### Database Technologies@ Alibaba Group

This year, the Data Analytics and Intelligence lab and Database Business Group of Alibaba Cloud Intelligence at will share some of the latest developments in Alibaba Group' database technologies and products during VLDB 2019.

The Data Analytics and Intelligence Lab at Alibaba Group is committed to the research and development of next-generation systems and algorithms for the distributed storage, data management, query processing, analytics, and machine learning on massive and heterogeneous data. The lab aims to provide efficient algorithm support and secure, reliable, and powerful computing engines for various industries and scenarios.

Database Business Group of Alibaba Cloud Intelligence supports all data processing applications of Alibaba Cloud and Alibaba Group. They also constantly investigate into new cloud native technologies to provide a full stack of database products and services, including Next generation cloud native OLTP database POLARDB, OLAP database AnalyticDB, Time series database: TSDB, Relational Database Services (RDS), non-relational database services (NoSQL), database utilities and tools such as data transfer service (DTS) and data backup service (DBS), as well as intelligence database tuning (Database Advisor).

### Our Accepted Papers@ VLDB 2019

#### Research Track

Ocean Vista: Gossip-Based Visibility Control for Speedy Geo-Distributed Transactions  
Dynamic Scaling for Parallel Graph Computations  
DPSAaS: MultiDimensional Data Sharing and Analytics as Services under Local Differential Privacy  
Yugong: Geo-Distributed Data and Job Placement at Scale

#### Industrial Track

AliGraph: A Comprehensive Graph Neural Network Platform  
AnalyticDB: Realtime OLAP Database System at Alibaba Cloud

### Our Highlights@ VLDB 2019

August 27<sup>th</sup> Invited Industrial Talk

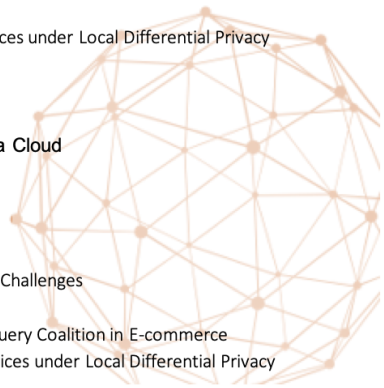
Feifei Li, Distinguished Engineer at Alibaba Group

Cloud Native Database Systems at Alibaba: Opportunities and Challenges

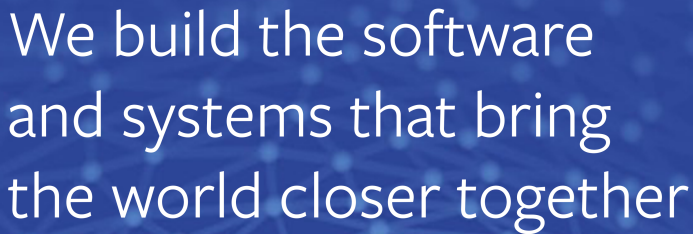
August 29<sup>th</sup> Demonstrations

SHOAL: Large-scale Hierarchical Taxonomy via Graph-based Query Coalition in E-commerce

DPSAaS: Multi-Dimensional Data Sharing and Analytics as Services under Local Differential Privacy



Facebook is pleased to be a platinum sponsor of VLDB 2019



We build the software  
and systems that bring  
the world closer together

Facebook researchers and engineers are constant innovators as they design and build next generation, scalable, fast, reliable, and efficient systems.

From distributed systems to data centers, hardware, storage, mobile, and beyond, the Facebook platform is our lab for research, development, and innovation.

To learn more about research at Facebook, visit us at our booth and at [research.fb.com](https://research.fb.com)

**facebook** research



## Welcome to join Huawei Database Team

### Research Direction >>

- Distributed OLAP/OLTP/in-memory
- Cloud data management
- Hardware acceleration: GPU/FPGA, RDMA, NVRAM
- Data management solution for IoT/mobile devices

### What you can expect >>

- End to end impact: opportunity to perform research prototyping, followed by architecture design and technology transfer into real products with support from HQ team
- State of the art R&D: strong collaboration with top universities in the globe, publish on top tier conferences and apply for patents
- Technical leadership: serve as architect and tech lead in production and mentor junior team members

### Contact US >>

Interested? Contact our Recruiting Manager,

Yi Gui:

Tel: +86 18729032515

Email: [guiyi@huawei.com](mailto:guiyi@huawei.com)



## IBM Research



## IBM Research is hiring

We live in a moment of remarkable change and opportunity. Data and technology are transforming industries and societies, ushering in a new era of Artificial Intelligence. IBM Research is a leader in this worldwide transformation, building on a long history of innovation.

For more than seven decades, IBM Research has defined the future of technology. Our scientists, among them six Nobel Laureates and six Turing Award winners, have produced ten U.S. National Medals of Technology and five U.S. National Medals of Science. Along the way we helped put a man on the moon, defeated Kasparov at chess, and built a Jeopardy! champion named Watson.

At IBM, you can achieve what others think is impossible. And in doing so, you'll play a significant role in shaping the future. Join us.

**Discover what you can do at IBM Research**

[ibm.com/jobs](http://ibm.com/jobs)



© Copyright IBM Corporation 2017. IBM, the IBM logo and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. See current list at [ibm.com/trademark](http://ibm.com/trademark). Other product and service names might be trademarks of IBM or other companies.

```
connectionString (
    uriParseExp(listenerNameString, "", port) =>
    listenerName = ListenerName.normalised(listenerNameString, host, port) =>
    EndPoint(host, port.toInt, listenerName, securityProtocol) =>
    EndPoint(host, port.toInt, listenerName, securityProtocol) =>
    => KafkaException(s"Unable to parse $listenerNameString")
)
```

# Data at **LinkedIn**

Extremely rich datasets driven by our **630M+** global members in over **200 countries and territories** plus **billions of nodes** in our **economic graph** form the back bone of our products.

Join our team of **AI experts, data scientists, infrastructure gurus, and dedicated researchers** helping connect the world's professionals to make them more productive and successful.

## Open roles within Data Engineering at LinkedIn include:

- Data Mining / Machine Learning
- Systems & Infrastructure
- Data Science



**Interested in learning more?** Come ask our team members about Engineering openings at our **VLDB 2019 Sponsor Booth, 2<sup>nd</sup> Fl., California Foyer!**

[careers.linkedin.com](https://careers.linkedin.com)

[data.linkedin.com](https://data.linkedin.com)





# Microsoft Research

Where leading scientists and engineers tackle complex problems to improve lives.

[aka.ms/VLDB-2019](https://aka.ms/VLDB-2019)

---

# GIANT IDEAS

**We are fearless in  
our innovation.**

Think what we can do for your career.

We're seeking great minds like you to help us build the next generation in database technology. Join the team that is helping the world's most sophisticated organizations transform their industries by harnessing the power of data.

 [www.mongodb.com/careers](http://www.mongodb.com/careers)



 [www.facebook.com/MongoDB](https://www.facebook.com/MongoDB)

 [@MongoDB](https://twitter.com/MongoDB) | [@MongoDBCareers](https://twitter.com/MongoDBCareers)

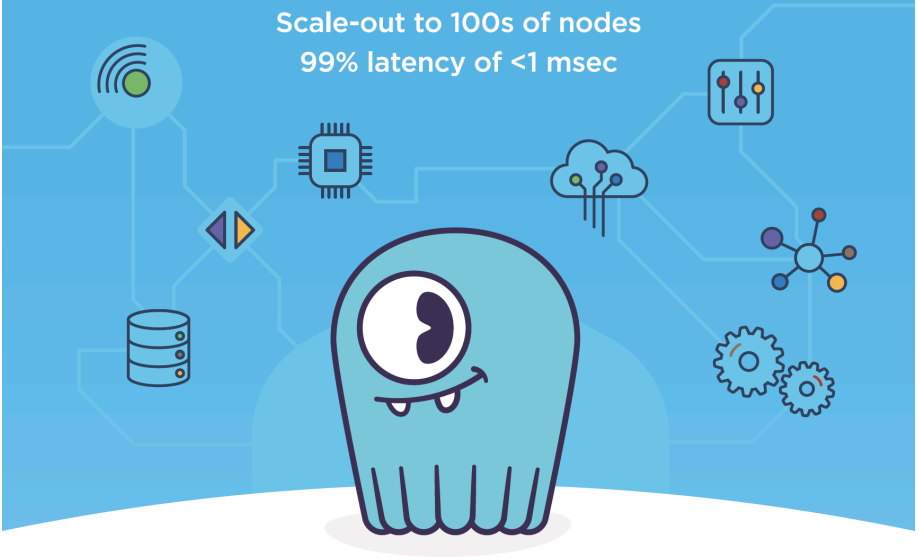
# SCYLLA.

## The Real-Time Big Data Database

Scale-up to 1,000,000 IOPS per node

Scale-out to 100s of nodes

99% latency of <1 msec



Take Scylla on a 1-hour Test Drive:  
[scylladb.com/test-drive](https://scylladb.com/test-drive)



# THE DATA WAREHOUSE BUILT FOR THE CLOUD



**Enabling every organization with the power, flexibility and instant elasticity to be data-driven.**

## AT SNOWFLAKE WE ARE:

Passionate, Innovative,  
Ambitious

*We challenge ourselves at Snowflake to rethink what's possible for the data warehouse and deliver on that. We're looking for people who share that same passion and ambition.*

Snowflake is hiring top database developers from around the world to grow our teams in San Mateo, Seattle, and our newly formed European Database R&D Center located in Berlin.

Visit us at :  
[www.snowflake.com/careers](http://www.snowflake.com/careers)

See yourself at  + a b l e a u<sup>®</sup>

**A mission that matters:**

We help people see and understand data.

Seven words that drive everything we do. And they've never been more important.

Our data visualization products are transforming the way people use data to solve problems. We make analyzing data fast and easy, beautiful and useful.

It's software for anyone and everyone.

**We are Tableau.**



# TDSQL

TDSQL is a distributed DBMS developed in Tencent, which features high availability, global deployment, horizontal scalability, high performance and enterprise-grade security. TDSQL is also a comprehensive database solution with facilities like smart database administration and automatic monitoring. These facilities not only help eliminate human error, but also make TDSQL cost less to run, perform better and more reliable.

## Global Deployment

TDSQL supports multi-datacenter and multi-region deployments globally.

## Horizontal Scalability

Optimized auto-sharding mechanism realizes seamless online scaling-out, thus ensuring linear increase of cluster performance and capacity.

## HTAP Capabilities

Distributed transaction and query processing prepare TDSQL for both OLTP and OLAP workloads.

## Enterprise-grade Security

TDSQL supports data encryption, database firewall and SSL connection, to achieve full-lifecycle data protection.

## Strong Consistency

Improved strong consistency. The performance of TDSQL's cross-data-center synchronization is comparable to that of asynchronous replication.

## Autonomous Database

Chitu and BianQue, two auxiliary systems in TDSQL solution, provide services like monitoring, data backup & restore, intelligent trouble-shooting and system diagnosis.

## Our customers



Follow us to learn more  
TDSQL@tencent.com

<http://cloud.tencent.com/product/dcdb>



**THE ONLY SCALABLE GRAPH DATABASE FOR THE ENTERPRISE**

**Unleash the power of interconnected data  
for deeper insights and better outcomes**

Through its Native Parallel Graph™ technology, the TigerGraph™ graph platform represents what's next in the graph database evolution: a complete, distributed, parallel graph computing platform supporting web-scale data analytics in real-time. Combining the best ideas (MapReduce, Massively Parallel Processing, and fast data compression/decompression) with fresh development, TigerGraph delivers what you've been waiting for: the speed, scalability, and deep exploration/querying capability to extract more business value from your data.

**Try the Developer Edition Now:**

**[WWW.TIGERGRAPH.COM/DOWNLOAD](http://WWW.TIGERGRAPH.COM/DOWNLOAD)**



X u p e r C h a i n

# Baidu XUPER

Independent Research and High-performance Blockchain Solution

## XuperChain Network

XuperChain has a strong network, high concurrency, effective verification ability and scalable storage capacity, which can protect your business in all directions.



## XUPER Product Family



## XuperChain Technical Characteristics

Supernode Technology	Inner-Chain Parallel Execution	Pluggable Consensus
3D Network	Smart Contracts	Network Latency

Website: <https://xchain.baidu.com/>



The eBay logo is centered in a teal square. The word "eBay" is written in a white, lowercase, sans-serif font. The background of the entire slide is a grid of four squares: a light blue square in the top-left, a dark blue square in the top-right, a teal square in the bottom-left (containing the logo), and a dark blue square in the bottom-right.

Marketplace: [www.ebay.com](http://www.ebay.com)

Company info: [www.ebayinc.com](http://www.ebayinc.com)





## Megagon Labs

Megagon Labs (formerly Recruit Institute of Technology) is the technology hub and research lab for Recruit Holdings, a company that provides over 200 online services in the areas of human resources, travel, housing, education, restaurants and many other areas in which people make daily lifestyle decisions.

We conduct world-class research in the several areas, including data management, data integration, natural language processing, machine learning, and affective computing. We collaborate with research in universities and publish in top-notch conferences.

Megagon Labs is hiring Research Scientists and Software Engineers. Visit our website at [www.megagon.ai](http://www.megagon.ai) for more information.

### 2019 VLDB Women in Database Research Award



Congratulations!  
Wang-Chiew Tan, Head of Research, received the award for her contributions to data provenance and integration spanning theory and practice.

Conference Awards Session  
August 29 (Thursday) 9:00-10:30  
at Room: Sacramento/San Francisco

### Paper and Poster Presentations by Megagon Labs

#### Subjective Databases

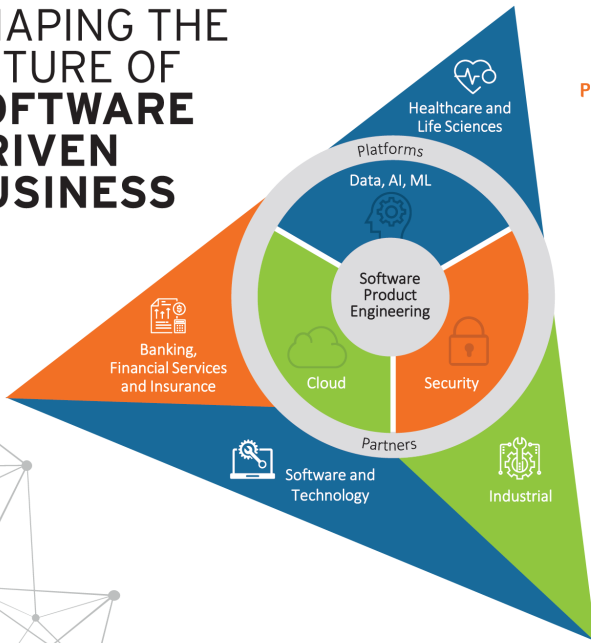
Research Session 20: Usability and data mining  
August 29 (Thursday) 11:00-12:30 at Emerald Bay room

Poster session:  
August 29 (Thursday) 17:30-18:30 at Room: Jogging Track

---

[www.megagon.ai](http://www.megagon.ai)

# SHAPING THE FUTURE OF SOFTWARE DRIVEN BUSINESS



Publicly listed in  
**BSE & NSE:**  
**PERSISTENT**



Revenue  
**\$480.97M**  
for FY 2018-2019

Presence in  
**15+**  countries



**10000+**  
employees



**25+** years of bringing software to life and life to software

[info@persistent.com](mailto:info@persistent.com)

[www.persistent.com](http://www.persistent.com)

© 2019 Persistent Systems Ltd. All rights reserved.

# Join the **SAP HANA** Database Campus!

Are you passionate about Software Development and working on your IT degree? Do you want to start your career by working on the next generation database technologies?

Then join us – The **SAP HANA Database Campus** team!

The **SAP HANA** team develops a platform that performs parallel in-memory processing of huge data sets to offer extremely fast real-time responses for analytic and transactional queries. The platform also provides libraries for predictive, planning, text processing, spatial, and business analytics – all on the same architecture.

Learn more about the **SAP HANA Database Campus** here:





# A NEW PARADIGM for Cloud Analytics



Most Innovative Product  
**WINNER**



## The Power of SQL without Code

- ▶ Visual Query Environment
- ▶ Speed with no data limits
- ▶ Semi-structured data support

## Modern Cloud-Native Technology

- ▶ SQL Compiler for Nested Data
- ▶ Multi-Tenant Distributed System
- ▶ Built with Rust, Go, and GraphQL

**WE'RE HIRING. JOIN OUR TEAM.**

[sigma.bi/careers](https://sigma.bi/careers)





## PLATINUM SPONSORS



## GOLD SPONSORS



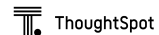
Megagon Labs



PERSISTENT



## SILVER SPONSORS



## BRONZE SPONSORS



## EXHIBITORS

