

41st
International
Conference on

Very Large Data Bases

Kohala Coast, Hawai`i, USA

August 31 - September 4, 2015



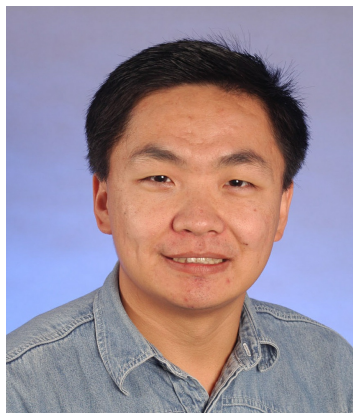
Welcome Message

VLDB is a premier annual international forum for data management and database researchers, vendors, practitioners, application developers, and users. The annual conference consists of a mix of research talks, tutorials, demonstrations, and workshops. Its topical coverage includes current issues in data management, database, and information systems research. Data management and databases remain among the main technological cornerstones of the applications of the twenty-first century. With the emergence of Big Data, data-related technologies are becoming more important than ever before. VLDB 2015 is taking place at the Hilton Waikoloa Village on the beautiful Kohala Coast on the northwestern side of the Big Island of Hawai'i. Hawai'i Island is the largest of the Hawaiian Islands; it is also referred to as the "Orchid Island" and more famously defined by its size as the Big Island, making it an especially appropriate setting for a conference whose concerns include Big Data.

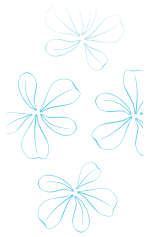
The VLDB 2015 conference received a large number of submissions to its various tracks. The research track received 710 high-quality submissions, with 139 papers accepted for presentation at the conference. In addition to those papers above, we also have 21 roll-over papers from VLDB 2014, resulting in a research program with 160 papers in total. The industrial track received 68 submissions, of which 20 were accepted. The demonstration track received 148 submissions, of which 49 were accepted. We are glad to see that the VLDB community is continuously developing new technologies in the space of data management to make an impact on businesses, the sciences, and society as a whole. We see a variety of topics ranging from traditional indexing, query processing and optimization, search, graph data management, to crowdsourcing and big data management. VLDB 2015 also features 6 tutorials, 4 keynote speeches, and 2 panels, a variety of exciting workshops, as well as the presentation of the 2015 Turing Award Lecture. The result for VLDB 2015 is an exciting technical program with many interesting presentations, including poster sessions featuring all of the research papers presented at the conference as well as papers from the VLDB Journal.

Michael Carey
VLDB 2015 General Chair

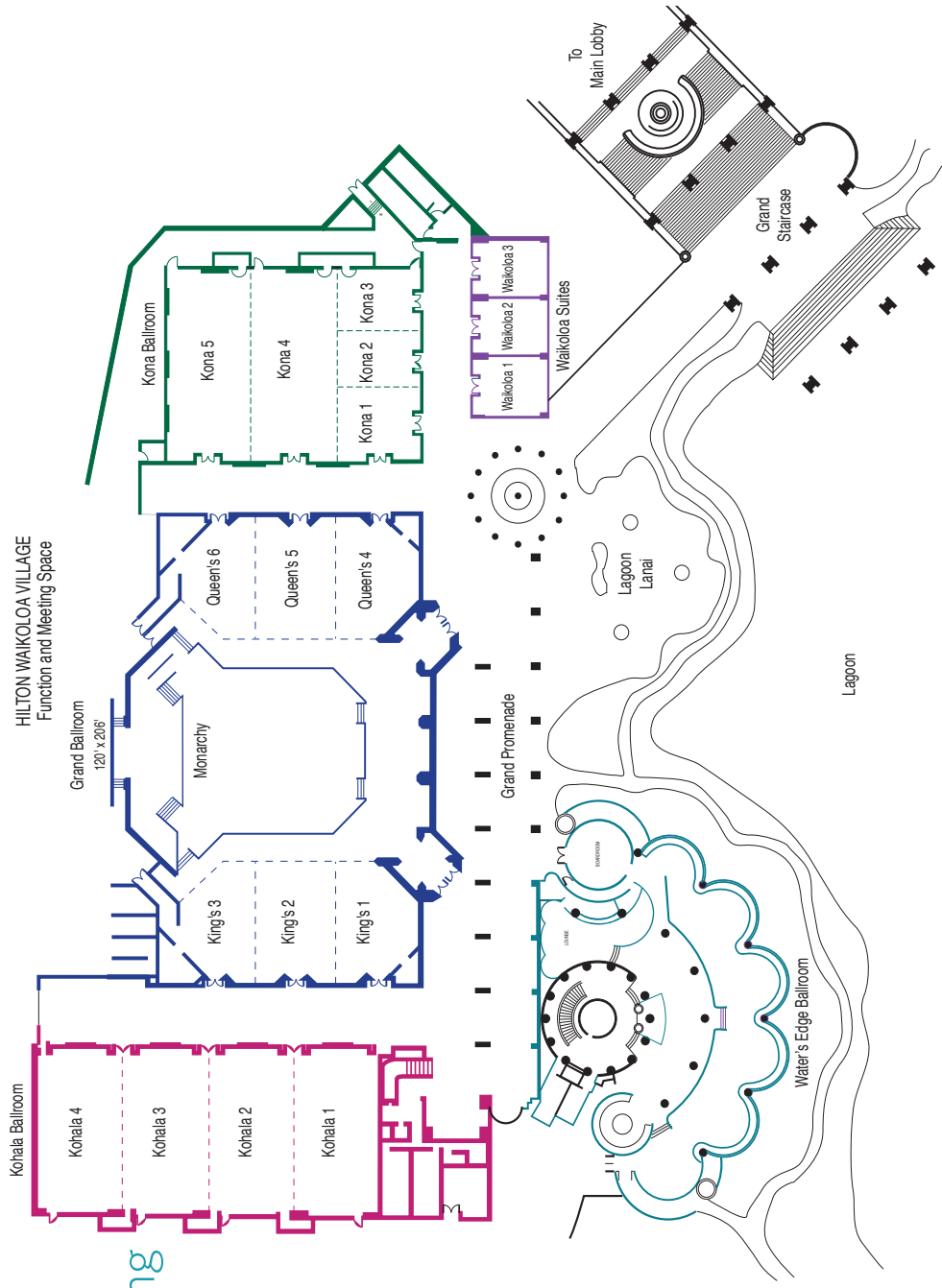
Chen Li and Volker Markl
VLDB 2015 Program Committee Chairs







Meeting Space



Sponsors

Platinum



Gold



Silver



Bronze



Exhibitors



Monday AT A GLANCE, AUG 31, 2015

Time	Kings 1	Kings 2	Kings 3	Queens 4	Queens 5	Queens 6	Kona 4	Kona 1-2-3
7:30-8:30	Light Breakfast (Grand Promenade)							
9:00-18:00	BIG-O(Q)	IMDM	ADMS	TPCTC	BIRTE	PhD workshop		
19:00-21:00	Welcome Buffet Reception (Lagoon Lanai)							

Tuesday AT A GLANCE, SEP 1, 2015

Time	Kings 1	Kings 2	Kings 3	Queens 4	Queens 5	Queens 6	Kona 4	Kona 1-2-3
7:30-8:30	Light Breakfast (Grand Promenade)							
8:30-10:15	Industrial Keynote: Juan Loaiza; Academic Keynote: Anastasia Ailamaki (Monarchy Ballroom)							
10:15-10:45	Coffee Break (Grand Promenade)							
10:30-12:15	Research 1: Big Data Systems Analysis	Research 2: Caching and Indexing	Industrial 1: Crowdsourcing, Data Cleaning, and Using Textual Knowledge Bases	Research 3: Data Mining	Research 4: Graph Mining	Tutorial 1: A Time Machine for Information: Looking Back to Look Forward	Demo 1: Data Mining, Graph, Text, and Semi-structured Data	
12:15-13:30	Lunch (Lagoon Lanai)							
13:30-15:00	Turing Award Lecture: Michael Stonebraker (Monarchy Ballroom)							
15:00-15:30	Coffee Break (Grand Promenade)							
15:30-17:00	Research 5: Graph Processing 1	Research 6: Information Integration	Industrial 2: Big Data Systems	Research 7: Query Interfaces and Languages	Research 8: Social Computing and Recommendations	Tutorial 2: On Uncertain Graphs Modeling and Queries	Demo 2: Information Retrieval, Data Quality, and Provenance	
17:15-19:00	Reception and Poster Session 1 (Kohala Ballroom)							

Wednesday AT A GLANCE, SEP 2, 2015

Time	Kings 1	Kings 2	Kings 3	Queens 4	Queens 5	Queens 6	Kona 4	Kona 1-2-3
7:30-8:30	Light Breakfast (Grand Promenade)							
8:30-10:00	Industrial Keynote: Todd Walter; Academic Keynote: Magdalena Balazinska (Monarchy Ballroom)							
10:00-10:30	Coffee Break (Grand Promenade)							
10:30-12:15	Research 9: Graph Processing 2	Research 10: Novel DB Architectures	Industrial 3: Real-time and Interactive Analytics	Research 11: Social Network Analysis	Research 12: Query Processing 1	Tutorial 3: Structured Analytics in Social Media	Demo 3: Systems, User Interfaces, and Visualization	
12:15-13.30	Lunch (Lagoon Lanai)							
13:30-15:00	Research 13: Graph Processing 3	Research 14: Novel Hardware Architectures	Industrial 4: Novel Approaches to Modern Data Processing	Research 15: Query Optimization	Research 16: Crowdsourcing and Social Network Analysis	Tutorial 4: SQL-on-Hadoop Systems (1/2)		Panel 1: 40-years VLDB
15:00-15:30	Coffee Break (Grand Promenade)							
15:30-17:00	Research 17: Graph Processing Systems	Research 18: Novel Hardware Architectures 2	Industrial 5: In-memory Data Management	Research 19: Social Network Analysis	Research 20: Ranking and Top-K	Tutorial 5: SQL-on-Hadoop Systems (2/2)	Demo 1: Data Mining, Graph, Text, and Semi-structured Data	
18:00-21:00	Banquet (Palace Lawn)							

Thursday AT A GLANCE, SEP 3, 2015

Time	Kings 1	Kings 2	Kings 3	Queens 4	Queens 5	Queens 6	Kona 4	Kona 1-2-3
7:30-8:30	Light Breakfast (Grand Promenade)							
8:30-10:00	Award Ceremony (Monarchy Ballroom)							
10:00-10:30	Coffee Break (Grand Promenade)							
10:30-12:00	Research 21: Spatial Databases	Research 22: Search	Industrial 6: Logging, Parallel Processing, and Graph Processing	Research 23: Logic and Semantics	Research 24: Innovative Systems	Tutorial 6: Truth Discovery and Crowdsourcing Aggregation: A Unified Perspective	Demo 2: Information Retrieval, Data Quality, and Provenance	
12:00-13:30	Lunch (Lagoon Lanai)							
13:30-15:00	Research 25: Probabilistic Data Processing and Approximation	Research 26: Query Processing 2	Industrial 7: Privacy and Visualization	Research 27: Data Warehousing, Search, and Ranking	Research 28: Novel DB Architectures, Novel Hardware, and Resource Management	Tutorial 7: Real Time Analytics: Algorithms and Systems (1/2)		Panel 2: Designing for Interaction: Broadening our View of Working with Data
15:00-15:30	Coffee Break (Grand Promenade)							
15:30-17:00	Research 29: Privacy and Security	Research 30: Logic Programming, Web Data Management, and Query Processing	Research 31: Structure and Dependency Discovery, and Spatial Databases	Research 32: Time Series and Streams	Research 33: Transaction Processing	Tutorial 8: Real Time Analytics: Algorithms and Systems (2/2)	Demo 3: Systems, User Interfaces, and Visualization	
17:15-19:00	Reception and Poster Session 2 (Kohala Ballroom)							

Friday AT A GLANCE, SEP 4, 2015

Time	Kings 1	Kings 2	Kings 3	Queens 4	Queens 5	Queens 6	Kona 4	Kona 1-2-3
7:30-8:30	Light Breakfast (Grand Promenade)							
9:00-18:00	DMAH	BPOE		BOSS				

Monday Aug 31st 08:00-09:00

Light Breakfast

Location: Grand Promenade

Monday Aug 31st 09:00-10:30

Big-O(Q) Session 1

Location: Kings 1

Workshop on Big-Graphs Online Querying (Big-O(Q) 2015)

Arijit Khan (ETH Zurich), Prasenjit Mitra (Qatar Computing Research Institute), Cong Yu (Google Research)

IMDM Session 1

Location: Kings 2

Third International Workshop on In-Memory Data Management and Analytics (IMDM 2015)

Justin Levandoski (Microsoft Research), Andy Pavlo (Carnegie Mellon University), Arun Jagatheesan (Samsung R&D Center), Thomas Neumann (Technische Universität München)

ADMS Session 1

Location: Kings 3

Sixth International Workshop on Accelerating Data Management Systems Using Modern Processor and Storage Architectures (ADMS 2015)

Rajesh Bordawekar (IBM T.J. Watson Research Center), Buğra Gedik (Bilkent University), Tirthankar Lahiri (Oracle), Christian A. Lang (Acelot Inc)

TPCTC Session 1

Location: Queens 4

Seventh TPC Technology Conference on Performance Evaluation & Benchmarking (TPCTC 2015)

Raghunath Nambiar (CISCO), Meikel Poess (Oracle)

BIRTE Session 1

Location: Queens 5

9th International Workshop on Business Intelligence for the Real Time Enterprise (BIRTE 2015)

Meichun Hsu (Hewlett-Packard), Malu Castellanos (Hewlett-Packard), Panos K Chrysanthis (University of Pittsburgh)

PhD Workshop Session 1

Location: Queens 6

PhD Workshop

Jennie Duggan (Northwestern University), Rachel Pottinger (University of British Columbia)

Monday Aug 31st 10:30-11:00

Refreshments: Coffee Break

Location: Grand Promenade

Monday Aug 31st 11:00-12:30

Big-O(Q) Session 2

Location: Kings 1

Workshop on Big-Graphs Online Querying (Big-O(Q) 2015)

Arijit Khan (ETH Zurich), Prasenjit Mitra (Qatar Computing Research Institute), Cong Yu (Google Research)

IMDM Session 2
Location: Kings 2

Third International Workshop on In-Memory Data Management and Analytics (IMDM 2015)

Justin Levandoski (Microsoft Research), Andy Pavlo (Carnegie Mellon University), Arun Jagatheesan (Samsung R&D Center), Thomas Neumann (Technische Universität München)

ADMS Session 2
Location: Kings 3

Sixth International Workshop on Accelerating Data Management Systems Using Modern Processor and Storage Architectures (ADMS 2015)

Rajesh Bordawekar (IBM T.J. Watson Research Center), Buğra Gedik (Bilkent University), Tirthankar Lahiri (Oracle), Christian A. Lang (Acelot Inc)

TPCTC Session 2
Location: Queens 4

Seventh TPC Technology Conference on Performance Evaluation & Benchmarking (TPCTC 2015)

Raghunath Nambiar (CISCO), Meikel Poess (Oracle)

BIRTE Session 2
Location: Queens 5

9th International Workshop on Business Intelligence for the Real Time Enterprise (BIRTE 2015)

Meichun Hsu (Hewlett-Packard), Malu Castellanos (Hewlett-Packard), Panos K Chrysanthis (University of Pittsburgh)

PhD Workshop Session 2
Location: Queens 6

PhD Workshop

Jennie Duggan (Northwestern University), Rachel Pottinger (University of British Columbia)

Monday Aug 31st 12:30-14:00

Lunch
Location: Lagoon Lanai

Monday Aug 31st 14:00-15:00

Big-O(Q) Session 3
Location: Kings 1

Workshop on Big-Graphs Online Querying (Big-O(Q) 2015)

Arijit Khan (ETH Zurich), Prasenjit Mitra (Qatar Computing Research Institute), Cong Yu (Google Research)

IMDM Session 3
Location: Kings 2

Third International Workshop on In-Memory Data Management and Analytics (IMDM 2015)

Justin Levandoski (Microsoft Research), Andy Pavlo (Carnegie Mellon University), Arun Jagatheesan (Samsung R&D Center), Thomas Neumann (Technische Universität München)

ADMS Session 3
Location: Kings 3

Sixth International Workshop on Accelerating Data Management Systems Using Modern Processor and Storage Architectures (ADMS 2015)

Rajesh Bordawekar (IBM T.J. Watson Research Center), Buğra Gedik (Bilkent University), Tirthankar Lahiri (Oracle), Christian A. Lang (Acelot Inc)

TPCTC Session 3

Location: Queens 4

Seventh TPC Technology Conference on Performance Evaluation & Benchmarking (TPCTC 2015)

Raghunath Nambiar (CISCO), Meikel Poess (Oracle)

BIRTE Session 3

Location: Queens 5

9th International Workshop on Business Intelligence for the Real Time Enterprise (BIRTE 2015)

Meichun Hsu (Hewlett-Packard), Malu Castellanos (Hewlett-Packard), Panos K Chrysanthis (University of Pittsburgh)

PhD Workshop Session 3

Location: Queens 6

PhD Workshop

Jennie Duggan (Northwestern University), Rachel Pottinger (University of British Columbia)

Monday Aug 31st 15:00-15:30

Refreshments: Coffee Break

Location: Grand Promenade

Monday Aug 31st 15:30-18:00

Big-O(Q) Session 4

Location: Kings 1

Workshop on Big-Graphs Online Querying (Big-O(Q) 2015)

Arijit Khan (ETH Zurich), Prasenjit Mitra (Qatar Computing Research Institute), Cong Yu (Google Research)

IMDM Session 4

Location: Kings 2

Third International Workshop on In-Memory Data Management and Analytics (IMDM 2015)

Justin Levandoski (Microsoft Research), Andy Pavlo (Carnegie Mellon University), Arun Jagatheesan (Samsung R&D Center), Thomas Neumann (Technische Universität München)

ADMS Session 4

Location: Kings 3

Sixth International Workshop on Accelerating Data Management Systems Using Modern Processor and Storage Architectures (ADMS 2015)

Rajesh Bordawekar (IBM T.J. Watson Research Center), Buğra Gedik (Bilkent University), Tirthankar Lahiri (Oracle), Christian A. Lang (Acelot Inc)

TPCTC Session 4

Location: Queens 4

Seventh TPC Technology Conference on Performance Evaluation & Benchmarking (TPCTC 2015)

Raghunath Nambiar (CISCO), Meikel Poess (Oracle)

BIRTE Session 4

Location: Queens 5

9th International Workshop on Business Intelligence for the Real Time Enterprise (BIRTE 2015)

Meichun Hsu (Hewlett-Packard), Malu Castellanos (Hewlett-Packard), Panos K Chrysanthis (University of Pittsburgh)

PhD Workshop Session 4

Location: Queens 6

PhD Workshop

Jennie Duggan (Northwestern University), Rachel Pottinger (University of British Columbia)

Monday Aug 31st 19:00-21:00

Opening Buffet Reception

Location: Lagoon Lanai

Chair: Sam Madden

Tuesday Sep 1st 07:30-08:30

Light Breakfast

Location: Grand Promenade

Tuesday Sep 1st 08:30-10:15

Industrial Keynote: Juan Loaiza; Academic Keynote: Anastasia Ailamaki

Location: Monarchy Ballroom

Chair: Volker Markl

Engineering Database Hardware and Software Together

Juan Loaiza, Oracle



Bio: As Senior Vice President of Systems Technology at Oracle, Juan Loaiza is in charge of developing the mission-critical capabilities of Oracle Database, including data and transaction management, high availability, performance, in-memory processing, enterprise replication, and Oracle Exadata. Mr. Loaiza joined the Oracle Database development organization in 1988. Mr. Loaiza holds BS and MS degrees in computer science from the Massachusetts Institute of Technology.

Databases and Hardware: The Beginning and Sequel of a Beautiful Friendship

Anastasia Ailamaki, EPFL



Bio: Anastasia Ailamaki is a Professor of Computer and Communication Sciences at the Ecole Polytechnique Federale de Lausanne (EPFL) in Switzerland. Her research interests are in data-intensive systems and applications, and in particular (a) in strengthening the interaction between the database software and emerging hardware and I/O devices, and (b) in nevermating data management to support computationally-demanding, data-intensive scientific applications. She has received an ERC Consolidator Award (2013), a Finmeccanica endowed chair from the Computer Science Department at Carnegie Mellon (2007), a European Young Investigator Award from the European Science Foundation (2007), an Alfred P. Sloan Research Fellowship (2005), eight best-paper awards in database, storage, and computer architecture conferences (2001-2012), and an NSF CAREER award (2002). She holds a Ph.D. in Computer Science from the University of Wisconsin-Madison in 2000. She is the vice chair of the ACM SIGMOD community, a senior member of the IEEE, and has served as a CRA-W mentor. She is a member of the Global Agenda Council for Data, Society and Development of the World Economic Forum.

Tuesday Sep 1st 10:15-10:45

Refreshments: Coffee Break

Location: Grand Promenade

Research 1: Big Data Systems Analysis

Location: Kings 1

Chair: Philippe Cudre-Mauroux

Shared Execution of Recurring Workloads in MapReduce

Chuan Lei (Worcester Polytechnic Institute), Zhongfang Zhuang (Worcester Polytechnic Institute), Elke Rundensteiner (Worcester Polytechnic Institute (US)), Mohamed Eltabakh (Worcester Polytechnic Institute)

A Performance Study of Big Data on Small Nodes

Dumitrel Loghin (NUS), Bogdan Tudor (NUS), Hao Zhang (NUS), Beng Chin Ooi (NUS), Yong Meng Teo (NUS)

Understanding the Causes of Consistency Anomalies in Apache Cassandra

Hua Fan (University of Waterloo), Aditya Ramaraju (University of Waterloo), Marlon McKenzie (University of Waterloo), Wojciech Golab (University of Waterloo), Bernard Wong (University of Waterloo)

Compaction management in distributed key-value datastores

Muhammad Yousuf Ahmad (McGill University), Bettina Kemme (McGill)

Fuzzy Joins in MapReduce: An Experimental Study

Ben Kimmitt (University of Victoria), Venkatesh Srinivasan (University of Victoria), Alex Thomo (University of Victoria)

Research 2: Caching and Indexing

Location: Kings 2

Chair: Jens Dittrich

Sharing Buffer Pool Memory in Multi-Tenant Relational Database-as-a-Service

Vivek Narasayya (Microsoft), Ishai Menache (Microsoft Research), Mohit Singh (Microsoft Research), Feng Li (Microsoft Research), Manoj Syamala (Microsoft Research), Surajit Chaudhuri (Microsoft)

Optimal Probabilistic Cache Stampede Prevention

Andrea Vattani (Goodreads Amazon Inc), Flavio Chierichetti (Sapienza), Keegan Lowenstein (Bugsnag Inc.)

Indexing Highly Dynamic Hierarchical Data

Jan Finis (TU München), Robert Brunel (Technische Universität München), Alfons Kemper (TU München), Thomas Neumann (TU München), Norman May (SAP AG), Franz Faerber (SAP SE)

BF-Tree: Approximate Tree Indexing

Manos Athanassoulis (EPFL), Anastasia Ailamaki (EPFL)

Industrial 1: Crowdsourcing, Data Cleaning, and Using Textual Knowledge Bases

Location: Kings 3

Chair: Kyuseok Shim

Argonaut: Macrotask Crowdsourcing for Complex Data Processing

Daniel Haas (UC Berkeley), Jason Ansel (GoDaddy), Lydia Gu (GoDaddy), Adam Marcus (Unlimited Labs)

FIT to monitor feed quality

Tamraparni Dasu (AT&T Labs-Research), Vladislav Shkapenyuk (AT&T Labs-Research), Divesh Srivastava (AT&T Labs-Research), Deborah Swayne (AT&T Labs-Research)

ConfSeer: Leveraging Customer Support Knowledge Bases for Automated Misconfiguration Detection

Rahul Potharaju (Microsoft), Joseph Chan (Microsoft), Luhui Hu (Microsoft), Cristina Nita-Rotaru (Purdue University), Mingshi Wang (Microsoft), Liyan Zhang (Microsoft), Navendu Jain (Microsoft Research)

Research 3: Data Mining

Location: Queens 4

Chair: Mourad Ouzzani

SRS: Solving c-Approximate Nearest Neighbor Queries in High Dimensional Euclidean Space with a Tiny Index

Yifang Sun (University of New South Wales), Wei Wang (University of New South Wales), Jianbin Qin (University of New South Wales), Ying Zhang (University of Technology Sydney), Xuemin Lin (University of New South Wales)

Rare Time Series Motif Discovery from Unbounded Streams

Nurjahan Begum (UC Riverside), Eamonn Keogh (UC Riverside)

Beyond Itemsets: Mining Frequent Featuresets over Structured Items

Saravanan Thirumuruganathan (University of Texas at Arlington), Habibur Rahman (University of Texas at Arlington), Sofiane Abbar (Qatar Computing Research Institute), Gautam Das (University of Texas at Arlington)

Mining Revenue-Maximizing Bundling Configuration

Loc Do (Singapore Management University), Hady W. Lauw (Singapore Management Univ), Ke Wang (Simon Fraser University)

ALID: Scalable Dominant Cluster Detection

Lingyang Chu (Institute of Computing Technology UCAS), Shuhui Wang (Institute of Computing Technology UCAS), Siyuan Liu (Carnegie Mellon University), Qingming Huang (University of Chinese Academy of Sciences), Jian Pei (Simon Fraser University)

Research 4: Graph Mining

Location: Queens 5

Chair: Kamal Karlapalem

Leveraging Graph Dimensions in Online Graph Search

Yuanyuan Zhu (Wuhan University), Jeffrey Xu Yu (The Chinese University of Hong Kong), Lu Qin (University of Technology (Sydney))

Event Pattern Matching over Graph Streams

Chunyao Song (University of Massachusetts (Lowell)), Tingjian Ge (University of Massachusetts (Lowell)), Cindy Chen (University of Massachusetts (Lowell)), Jie Wang (University of Massachusetts (Lowell))

An Efficient Similarity Search Framework for SimRank over Large Dynamic Graphs

Yingxia Shao (Peking University), Bin Cui (Peking University), Lei Chen (Hong Kong University of Science and Technology), Mingming Liu (Peking University), Xing Xie (Microsoft Research)

Growing a Graph Matching from a Handful of Seeds

Ehsan Kazemi (EPFL), Seyed Hamed Hassani (ETHZ), Matthias Grossglauser (EPFL)

Association Rules with Graph Patterns

Wenfei Fan (University Edinburgh and Beihang University), Xin Wang (Southwest Jiaotong University), Yinghui Wu (Washington State University), Jingbo Xu, University Edinburgh and Beihang University)

Tutorial 1: A Time Machine for Information: Looking Back to Look Forward

Location: Queens 6

A Time Machine for Information: Looking Back to Look Forward

Xin Luna Dong, Wang-Chiew Tan

Demo 1: Data Mining, Graph, Text, and Semi-structured Data

Location: Kona 4

Chair: Tianli Mo

Evaluating SPARQL Queries on Massive RDF Datasets

Razen Harbi (King Abdullah University of Science and Technology), Ibrahim Abdelaziz (King Abdullah University of Science and Technology), Panos Kalnis (King Abdullah University of Science and Technology), Nikos Mamoulis (University of Ioannina)

Demonstration of Santoku: Optimizing Machine Learning over Normalized Data

Arun Kumar (University of Wisconsin-Madison), Mona Jalal (University of Wisconsin-Madison), Boqun Yan (University of Wisconsin-Madison), Jeffrey Naughton (University Wisconsin-Madison), Jignesh Patel (University of Wisconsin-Madison)

PRISM: Concept-preserving Summarization of Top-K Social Image Search Results

Boon-Siew Seah (Nanyang Technological University), Sourav S Bhowmick (Nanyang Technological University), Aixin Sun (Nanyang Technological University)

SPARTEX: A Vertex-Centric Framework for RDF Data Analytics

Ibrahim Abdelaziz (King Abdullah University of Science and Technology), Razen Harbi (King Abdullah University of Science and Technology), Semih Salihoglu (Stanford University), Panos Kalnis (King Abdullah University of Science and Technology), Nikos Mamoulis (University of Ioannina)

I2RS: A Distributed Geo-Textual Image Retrieval and Recommendation System

Lu Chen (Zhejiang University), Yunjun Gao (Zhejiang University), Zhihao Xing (Zhejiang University), Christian Jensen (Aalborg University), Gang Chen (Zhejiang University)

Reformulation-based query answering in RDF: alternatives and performance

Damian Bursztyn (INRIA), Francois Goasdoue (University of Rennes 1), Ioana Manolescu (INRIA)

TreeScope: Finding Structural Anomalies In Semi-Structured Data

Shanshan Ying (ADSC), Flip Korn, Barna Saha (University of Massachusetts Amherst), Divesh Srivastava (AT&T Labs-Research)

PERSEUS: An Interactive Large-Scale Graph Mining and Visualization Tool

Danai Koutra (Carnegie Mellon University), Di Jin (Carnegie Mellon University), Yuanchi Ning (Uber Technologies Inc.), Christos Faloutsos (Carnegie Mellon University)

Virtual eXist-db: Liberating Hierarchical Queries from the Shackles of Access Path Dependence

Curtis Dyreson (Utah State University), Sourav S Bhowmick (Nanyang Technological University), Ryan Grapp (Utah State University)

FLORIN - A System to Support (Near) Real-Time Applications on User Generated Content on Daily News

Qingyuan Liu (Temple University), Eduard Dragut (Temple University), Arjun Mukherjee (University of Houston), Weiyi Meng (Binghamton University)

A Framework for Clustering Uncertain Data

Erich Schubert (Ludwig-Maximilians-Universität München), Alexander Koos (Ludwig-Maximilians-Universität München), Tobias Emrich (Ludwig-Maximilians-Universität München), Andreas Züfle (Ludwig-Maximilians-Universität München), Klaus Schmid (Ludwig-Maximilians-Universität München), Arthur Zimek (Ludwig-Maximilians-Universität München)

Query-oriented summarization of RDF graphs

Sejla Cebiric (INRIA), Francois Goasdoue (University of Rennes 1), Ioana Manolescu (INRIA)

Universal-DB: Towards Representation Independent Graph Analytics

Yodsawalai Chodpathumwan (University of Illinois), Amirhossein Aleyassin (University of Illinois), Arash Termehchy (Oregon State University), Yizhou Sun (Northeastern University)

Tornado: A Distributed Spatio-Textual Stream Processing System

Ahmed Mahmood (Purdue University), Ahmed Aly (Purdue University), Thamir Qadah (Purdue University), El Kindi Rezig (Purdue University), Anas Daghistani (Purdue University), Amgad Madkour (Purdue University), Ahmed Abdelhamid (Purdue University), Mohamed Hassan (Purdue University), Walid Aref (Purdue University (USA)), Saleh Basalamah (Umm Al-Qura University)

S+EPP: Construct and Explore Bisimulation Summaries, plus Optimize Navigational Queries; all on Existing SPARQL Systems

Mariano Consens (University of Toronto), Valeria Fionda (University of Calabria), Shahan Khatchadourian (University of Toronto), Giuseppe Pirrò (ICAR-CNR)

GraphGen: Exploring Interesting Graphs in Relational Data

Konstantinos Xirogiannopoulos (University of Maryland at College Park), Udayan Khurana (University of Maryland at College Park), Amol Deshpande (University of Maryland at College Park)

StarDB: A Large-Scale DBMS for Strings

Majed Sahli (King Abdullah University of Science and Technology), Essam Mansour (QCRI), Panos Kalnis (King Abdullah University of Science and Technology)

Tuesday Sep 1st 12:15-13:30

Lunch

Location: Lagoon Lanai

Tuesday Sep 1st 13:30-15:00

Turing Award Lecture: Michael Stonebraker

Location: Monarchy Ballroom

Chair: David DeWitt

The Land Sharks are on the Squawk Box (How Riding a Bicycle across America and Building Postgres Have a Lot in Common)

Michael Stonebraker, MIT



Bio: Michael Stonebraker is an Adjunct Professor of Computer Science at MIT and recipient of the 2014 A.M. Turing Award from the ACM for his fundamental contributions to the concepts and practices underlying modern database systems. He specializes in database management systems and data integration, and has been a pioneer of database research and technology for more than 40 years. He is the author of scores of papers in this area. He was the main architect of the INGRES relational DBMS, the object-relational DBMS POSTGRES, and the federated data system, Mariposa; and principal architect of the C-Store column store database, H-Store main-memory OLTP engine, and SciDB array engine. He has started nine start-up companies to commercialize these database technologies and, more recently, Big Data technologies (Vertica, VoltDB, Paradigm4, Tamr). He is a member of the National Academy of Engineering and the American Academy of Arts and Sciences.

Tuesday Sep 1st 15:00-15:30

Refreshments: Coffee Break
Location: Grand Promenade

Tuesday Sep 1st 15:30-17:00

Research 5: Graph Processing 1
Location: Kings 1
Chair: Wook-Shin Han

Efficient Top-K SimRank-based Similarity Join

Wenbo Tao (Tsinghua University), Minghe Yu (Tsinghua University), Guoliang Li (Tsinghua University)

MOCgraph: Scalable Distributed Graph Processing Using Message Online Computing

Chang Zhou (Peking University), Jun Gao (Peking University), Binbin Sun (Huawei), Jeffrey Xu Yu (The Chinese University of Hong Kong)

The More the Merrier: Efficient Multi-Source Graph Traversal

Manuel Then (Technische Universität München), Moritz Kaufmann (Technische Universität München), Fernando Chirigati (New York University), Tuan-Anh Hoang-Vu (New York University), Kien Pham (New York University), Alfons Kemper (TUM), Thomas Neumann (TU Munich (Germany)), Huy Vo (New York University)

Efficient Partial-Pairs SimRank Search on Large Networks

Weiren Yu (Imperial College London), Julie McCann (Imperial College London)

Exploiting Vertex Relationships in Speeding up Subgraph Isomorphism over Large Graphs

Xuguang Ren (Griffith University), Junhu Wang (Griffith University)

Research 6: Information Integration

Location: Kings 2
Chair: David Maier

Preference-aware Integration of Temporal Data

Bogdan Alexe (IBM Almaden Research Center), Mary Roth (UCSC and IBM Research), Wang-Chiew Tan (UCSC)

Optimizing the Chase: Scalable Data Integration under Constraints

George Konstantinidis (USC), Jose-Luis Ambite (USC)

Supervised Meta-blocking

George Papadakis (IMIS Research Center "Athena"), George Papastefanatos (IMIS Research Center "Athena"), Georgia Koutrika (HP Labs)

Enriching Data Imputation with Extensive Similarity Neighbors

Shaoxu Song (Tsinghua University), Aoqian Zhang (Tsinghua University), Lei Chen (Hong Kong University of Science and Technology), Jianmin Wang (Tsinghua University)

Industrial 2: Big Data Systems

Location: Kings 3
Chair: Guy Lohman

Gobblin: Unifying Data Ingestion for Hadoop

Lin Qiao (LinkedIn), Yanan Li (LinkedIn), Sahil Takiar (LinkedIn), Ziyang Liu (LinkedIn), Narasimha Veeramreddy (LinkedIn), Min Tu (LinkedIn), Ying Dai (LinkedIn), Issac Buenrostro (LinkedIn), Kapil Surlaker (LinkedIn), Shirshanka Das (LinkedIn), Chavdar Botev (LinkedIn)

Schema-Agnostic Indexing with Azure DocumentDB

Dharma Shukla (Microsoft), Shireesh Thota (Microsoft), Karthik Raman (Microsoft), Madhan Gajendran (Microsoft), Ankur Shah (Microsoft), Sergii Ziuzin (Microsoft), Krishnan Sundaram (Microsoft), Miguel Gonzalez Guajardo (Microsoft), Anna Wawrzyniak (Microsoft), Samer Boshra (Microsoft), Renato Ferreira (Microsoft), Mohamed Nassar (Microsoft), Michael Koltachev (Microsoft), Ji Huang (Microsoft), Sudipta Sengupta (Microsoft), Justin Levandoski (Microsoft), David Lomet (Microsoft)

Scaling Spark in the Real World

Michael Armbrust (Databricks Inc), Tathagata Das (Databricks Inc), Aaron Davidson (Databricks Inc), Ali Ghodsi (Databricks Inc), Andrew Or (Databricks Inc), Josh Rosen (Databricks Inc), Ion Stoica (UC Berkeley), Patrick Wendell (Databricks Inc), Reynold Xin (Databricks Inc), Matei Zaharia (MIT CSAIL)

Research 7: Query Interfaces and Languages

Location: Queens 4

Chair: Bernhard Mitschang

Answering Why-not Questions on Reverse Top-k Queries

Yunjun Gao (Zhejiang University), Qing Liu (Zhejiang University), Gang Chen (Zhejiang University), Baihua Zheng (Singapore Management University), Linlin Zhou (Zhejiang University)

SnapToQuery: Providing Interactive Feedback during Exploratory Query Specification

Lilong Jiang (Ohio State University), Arnab Nandi (Ohio State University)

Constructing an Interactive Natural Language Interface for Relational Databases

Fei Li (University of Michigan), H V Jagadish (University of Michigan)

A Natural Language Interface for Querying General and Individual Knowledge

Yael Amsterdamer (Tel Aviv University), Anna Kukliansky (Tel Aviv University), Tova Milo (Tel Aviv University)

Possible and Certain SQL Keys

Henning Köhler (Massey University), Sebastian Link (The University of Auckland), Xiaofang Zhou (The University of Queensland)

Research 8: Social Computing and Recommendations

Location: Queens 5

Chair: Mahashweta Das

D2P: Distance-Based Differential Privacy in Recommenders

Rachid Guerraoui (EPFL), Anne-Marie Kermarrec (INRIA), Rhicheek Patra (EPFL), Mahsa Taziki (EPFL)

Show Me the Money: Dynamic Recommendations for Revenue Maximization

Wei Lu (University of British Columbia), Shanshan Chen (University of British Columbia), Keqian Li (University of British Columbia), Laks V. S. Lakshmanan (University of British Columbia)

Finish Them!: Pricing Algorithms for Human Computation

Yihan Gao (UIUC), Aditya Parameswaran (Massachusetts Institute of Technology)

TransactiveDB: Tapping into Collective Human Memories

Michele Catasta (EPFL), Alberto Tonon (University of Fribourg), Djellel Eddine Difallah (University of Fribourg), Gianluca Demartini (eXascale Infolab), Karl Aberer (EPFL), Philippe Cudré-Mauroux (University of Fribourg)

Worker Skill Estimation in Team-Based Tasks

Habibur Rahman (University of Texas at Arlington), Saravanan Thirumuruganathan (University of Texas at Arlington), Senjuti Basu Roy (UW), Sihem Amer-Yahia (LIG), Gautam Das (University of Texas at Arlington)

Tutorial 2: On Uncertain Graphs Modeling and Queries

Location: Queens 6

On Uncertain Graphs Modeling and Queries

Arijit Khan, Lei Chen

Demo 2: Information Retrieval, Data Quality, and Provenance

Location: Kona 4

Chair: Tianli Mo

A Topic-based Reviewer Assignment System

Ngai Meng Kou (University of Macau), Leong Hou U (University of Macau), Nikos Mamoulis (University of Hong Kong), Yuhong Li (University of Macau), Ye Li (University of Macau), Zhiguo Gong (University of Macau)

Data Profiling with Metanome

Thorsten Papenbrock (Hasso-Plattner-Institute), Tanja Bergmann (Hasso-Plattner-Institute), Moritz Finke (Hasso-Plattner-Institute), Jakob Zwiener (Hasso-Plattner-Institute), Felix Naumann (Hasso-Plattner-Institute)

Provenance for SQL through Abstract Interpretation: Value-less, but Worthwhile

Tobias Müller (U Tübingen), Torsten Grust (U Tübingen)

SAASFEE: Scalable Scientific Workflow Execution Engine

Marc Bux (Humboldt-Universität zu Berlin), Jörgen Brandt (Humboldt-Universität zu Berlin), Carsten Lipka (Humboldt-Universität zu Berlin), Kamal Hakimzadeh (KTH Royal Institute of Technology), Jim Dowling (KTH Royal Institute of Technology), Ulf Leser (Humboldt Universität zu Berlin)

QOCO: A Query Oriented Data Cleaning System with Oracles

Moria Bergman (Tel Aviv University), Tova Milo (Tel Aviv University), Slava Novgorodov (Tel Aviv University), Wang-Chiew Tan (University of California Santa Cruz)

Collaborative Data Analytics with DataHub

Anant Bhardwaj (MIT), Amol Deshpande (University of Maryland), Aaron Elmore (University of Chicago), David Karger (MIT), Sam Madden (MIT), Aditya Parameswaran (University of Illinois at Urbana Champaign), Harihar Subramanyam (MIT), Eugene Wu (Columbia), Rebecca Zhang (MIT)

Mindtagger: A Demonstration of Data Labeling in Knowledge Base Construction

Jaeho Shin (Stanford University), Christopher Re (Stanford University), Mike Cafarella (University of Michigan)

Annotating Database Schemas to Help Enterprise Search

Eli Cortez (Microsoft), Philip Bernstein (Microsoft Research), Yeye He (Microsoft Research), Lev Novik (Microsoft)

KATARA: Reliable Data Cleaning with Knowledge Bases and Crowdsourcing

Xu Chu (University of Waterloo), John Morcos (University of Waterloo), Ihab Ilyas (University of Waterloo), Mourad Ouzzani (QCRI), Paolo Papotti (QCRI), Nan Tang (QCRI), Yin Ye (Google)

Gain Control over your Integration Evaluations Using iBench

Patricia Arocena (University of Toronto), Radu Ciucanu (University of Lille (INRIA), Boris Glavic (IIT), Renee Miller (University Toronto)

Janiform Intra-Document Analytics for Reproducible Research

Jens Dittrich (Saarland University), Patrick Bender (Saarland University)

EFQ: Why-Not Answer Polynomials in Action

Katerina Tzompanaki (Université Paris Sud), Nicole Bidoit (Université Paris Sud - INRIA), Melanie Herschel (University of Stuttgart)

Error Diagnosis and Data Profiling with Data X-Ray

Xiaolan Wang (University of Massachusetts Amherst), Mary Feng (University of Massachusetts Amherst and University of Iowa), Yue Wang (University of Massachusetts Amherst), Xin Luna Dong (Google Inc), Alexandra Meliou (University of Massachusetts Amherst)

Sharing and Reproducing Database Applications

Quan Pham (University of Chicago), Severin Thaler (University of Chicago), Tanu Malik (University of Chicago), Ian Foster (University of Chicago), Boris Glavic (IIT)

A Demonstration of TripleProv: Tracking and Querying Provenance over Web Data

Marcin Wylot (University of Fribourg), Philippe Cudré-Mauroux (University of Fribourg), Paul Groth (Elsevier Labs)

WADaR: Joint Wrapper and Data Repair

Stefano Ortona (University of Oxford), Giorgio Orsi (University of Oxford), Marcello Buoncristiano (Universita della Basilicata), Tim Furche (University of Oxford)

Wisteria: Nurturing Scalable Data Cleaning Infrastructure

Daniel Haas (UC Berkeley), Sanjay Krishnan (UC Berkeley), Jiannan Wang (UC Berkeley), Michael Franklin (UC Berkeley), Eugene Wu (Columbia University)

Reception and Poster Session 1

Location: Kohala Ballroom

Chair: Tianli Mo

Shared Execution of Recurring Workloads in MapReduce

Chuan Lei - Zhongfang Zhuang - Elke Rundensteiner - Mohamed Eltabakh

A Performance Study of Big Data on Small Nodes

Dumitrel Loghin - Bogdan Tudor - Hao Zhang - Beng Chin Ooi - Yong Meng Teo

Understanding the Causes of Consistency Anomalies in Apache Cassandra

Hua Fan - Aditya Ramaraju - Marlon McKenzie - Wojciech Golab - Bernard Wong

Compaction management in distributed key-value datastores

Muhammad Yousuf Ahmad - Bettina Kemme

Fuzzy Joins in MapReduce: An Experimental Study

Ben Kimmitt - Venkatesh Srinivasan - Alex Thomo

Sharing Buffer Pool Memory in Multi-Tenant Relational Database-as-a-Service

Vivek Narasayya - Ishai Menache - Mohit Singh - Feng Li - Manoj Syamala - Surajit Chaudhuri

Optimal Probabilistic Cache Stampede Prevention

Andrea Vattani - Flavio Chierichetti - Keegan Lowenstein

Indexing Highly Dynamic Hierarchical Data

Jan Finis - Robert Brunel - Alfons Kemper - Thomas Neumann - Norman May - Franz Faerber

BF-Tree: Approximate Tree Indexing

Manos Athanassoulis - Anastasia Ailamaki

SRS: Solving c-Approximate Nearest Neighbor Queries in High Dimensional Euclidean Space with a Tiny Index

Yifang Sun - Wei Wang - Jianbin Qin - Ying Zhang - Xuemin Lin

Rare Time Series Motif Discovery from Unbounded Streams

Nurjahan Begum - Eamonn Keogh

Beyond Itemsets: Mining Frequent Featuresets over Structured Items

Saravanan Thirumuruganathan - Habibur Rahman - Sofiane Abbar - Gautam Das

Mining Revenue-Maximizing Bundling Configuration

Loc Do - Hady W. Lauw - Ke Wang

ALID: Scalable Dominant Cluster Detection

Lingyang Chu - Shuhui Wang - Siyuan Liu - Qingming Huang - Jian Pei

Leveraging Graph Dimensions in Online Graph Search

Yuanyuan Zhu - Jeffrey Xu Yu - Lu Qin

Event Pattern Matching over Graph Streams

Chunyao Song - Tingjian Ge - Cindy Chen - Jie Wang

An Efficient Similarity Search Framework for SimRank over Large Dynamic Graphs

Yingxia Shao - Bin Cui - Lei Chen - Mingming Liu - Xing Xie

Growing a Graph Matching from a Handful of Seeds

Ehsan Kazemi - Seyed Hamed Hassani - Matthias Grossglauser

Association Rules with Graph Patterns

Wenfei Fan - Xin Wang - Yinghui Wu - Jingbo Xu

Efficient Top-K SimRank-based Similarity Join

Wenbo Tao - Minghe Yu - Guoliang Li

MOCgraph: Scalable Distributed Graph Processing Using Message Online Computing

Chang Zhou - Jun Gao - Binbin Sun - Jeffrey Xu Yu

The More the Merrier: Efficient Multi-Source Graph Traversal

Manuel Then - Moritz Kaufmann - Fernando Chirigati - Tuan-Anh Hoang-Vu - Kien Pham - Alfons Kemper - Thomas Neumann - Huy Vu

Efficient Partial-Pairs SimRank Search on Large Networks

Weiren Yu - Julie McCann

Exploiting Vertex Relationships in Speeding up Subgraph Isomorphism over Large Graphs

Xuguang Ren - Junhu Wang

Preference-aware Integration of Temporal Data

Bogdan Alexe - Mary Roth - Wang-Chiew Tan

Optimizing the Chase: Scalable Data Integration under Constraints

George Konstantinidis - Jose-Luis Ambite

Supervised Meta-blocking

George Papadakis - George Papastefanatos - Georgia Koutrika

Enriching Data Imputation with Extensive Similarity Neighbors

Shaoxu Song - Aoqian Zhang - Lei Chen - Jianmin Wang

Answering Why-not Questions on Reverse Top-k Queries

Yunjun Gao - Qing Liu - Gang Chen - Baihua Zheng - Linlin Zhou

SnapToQuery: Providing Interactive Feedback during Exploratory Query Specification

Lilong Jiang - Arnab Nandi

Constructing an Interactive Natural Language Interface for Relational Databases

Fei Li - H. V. Jagadish

A Natural Language Interface for Querying General and Individual Knowledge

Yael Amsterdamer - Anna Kukliansky - Tova Milo

Possible and Certain SQL Keys

Henning Kohler - Sebastian Link - Xiaofang Zhou

D2P: Distance-Based Differential Privacy in Recommenders

Rachid Guerraoui - Anne-Marie Kermarrec - Rhicheek Patra - Mahsa Taziki

Show Me the Money: Dynamic Recommendations for Revenue Maximization

Wei Lu - Shanshan Chen - Keqian Li - Laks V. S. Lakshmanan

Finish Them!: Pricing Algorithms for Human Computation

Yihan Gao - Aditya Parameswaran

TransactiveDB: Tapping into Collective Human Memories

Michele Catasta - Alberto Tonon - Djellel Eddine Difallah - Gianluca Demartini - Karl Aberer - Philippe Cudre-Mauroux

Worker Skill Estimation in Team-Based Tasks

Habibur Rahman - Saravanan Thirumuruganathan - Senjuti Basu Roy - Sihem Amer-Yahia - Gautam Das

Scalable Subgraph Enumeration in MapReduce

Longbin Lai - Lu Qin - Xuemin Lin - Lijun Chang

FrogWild! -- Fast PageRank Approximations on Graph Engines

Ioannis Mitliagkas - Michael Borokhovich - Alexandros Dimakis - Constantine Caramanis

Pregel Algorithms for Graph Connectivity Problems with Performance Guarantees

Da Yan - James Cheng - Kai Xing - Yi Lu - Wilfred Ng - Yingyi Bu

Blogel: A Block-Centric Framework for Distributed Computation on Real-World Graphs

Da Yan - James Cheng - Yi Lu - Wilfred Ng

LogGP: A Log-based Dynamic Graph Partitioning Method

Ning Xu - Lei Chen - Bin Cui

Coordination autoance in Database Systems

Peter Bailis - Alan Fekete - Michael Franklin - Ali Ghodsi - Joseph Hellerstein - Ion Stoica

A Scalable Search Engine for Mass Storage Smart Objects

Nicolas Ancaux - Saliha Lallali - Iulian Sandu Popa - Philippe Pucheral

Schema Management for Document Stores

Lanjun Wang - Oktie Hassanzadeh - Shuo Zhang - Juwei Shi - Limei Jiao - Jia Zou - Chen Wang

Supporting Scalable Analytics with Latency Constraints

Boduo Li - Yanlei Diao - Prashant Shenoy

Principles of Dataset Versioning: Exploring the Recreation/Storage Tradeoff

Souvik Bhattacharjee - Amit Chavan - Silu Huang - Amol Deshpande - Aditya Parameswaran

Inferring Continuous Dynamic Social Influence and Personal Preference for Temporal Behavior Prediction

Jun Zhang - Chaokun Wang - Jianmin Wang - Jeffrey Xu Yu

Influential Community Search in Large Networks

Rong-Hua LI - Lu Qin - Jeffrey Xu Yu - Rui Mao

Linearized and Single-Pass Belief Propagation

Wolfgang Gatterbauer - Stephan Gunnemann - Danai Koutra - Christos Faloutsos

Online Topic-Aware Influence Maximization

Shuo Chen - Ju Fan - Guoliang Li - Jianhua Feng - Kian-Lee Tan - Jinhui Tang

Walk, Not Wait: Faster Sampling Over Online Social Networks

Azade Nazi - Zhuojie Zhou - Saravanan Thirumuruganathan - Nan Zhang - Gautam Das

Work-Efficient Parallel Skyline Computation for the GPU

Kenneth Bogh - Sean Chester - Ira Assent

Memory-Efficient Hash Joins

R. Barber - G. Lohman - I. Pandis - V. Raman - R. Sidle - G. Attaluri - N. Chainani - S. Lightstone - D. Sharpe

MRCST: Compressing and Searching String Collections with Multiple References

Sebastian Wandelt - Ulf Leser

Trill: A High-Performance Incremental Query Processor for Diverse Analytics

Badrish Chandramouli - Jonathan Goldstein - Mike Barnett - Robert DeLine - John Platt - James Terwilliger - John Wernsing

Rapid Sampling for Visualizations with Ordering Guarantees

Albert Kim - Eric Blais - Aditya Parameswaran - Piotr Indyk - Sam Madden - Ronitt Rubinfeld

Argonaut: Macrotask Crowdsourcing for Complex Data Processing

Adam Marcus - Lydia Gu - Daniel Haas - Jason Ansel

FIT to monitor feed quality

Tamraparni Dasu - Vladislav Shkapenyuk - Divesh Srivastava - Deborah Swayne

ConfSeer: Leveraging Customer Support Knowledge Bases for Automated Misconfiguration Detection

Rahul Potharaju - Navendu Jain

Goblin: Unifying Data Ingestion for Hadoop

Lin Qiao - Kapil Surlaker - Shirshanka Das - Chavdar Botev - Yinan Li - Sahil Takiar - Henry Cai - Narasimha Veeramreddy - Min Tu - Ziyang Liu - Ying Dai

Schema-Agnostic Indexing with Azure DocumentDB

Dharma Shukla - Shireesh Thota - Karthik Raman - Madhan Gajendran - Ankur Shah - Sergii Ziuzin - Krishnan Sundaram - Anna Wawrzyniak - Samer Boshra - Mohamed Nassar - Michael Koltachev - Sudipta Sengupta - Justin Levandoski - David Lomet

Scaling Spark in the Real World

Michael Armbrust - Tathagata Das - Aaron Davidson - Ali Ghodsi - Andrew Or - Josh Rosen - Ion Stoica - Patrick Wendell - Reynold Xin - Matei Zaharia

JetScope: Reliable and Interactive Analytics at Cloud Scale

Eric Boutin - Jaliya Ekanayake - Anna Korsun - Jingren Zhou

Towards Scalable Real-time Analytics: An Architecture for Scale-out of OLxP Workloads

Jeffrey Pound - Anil Goel - Nathan Auch - Franz Faerber - Francis Gropengiesser - Christian Mathis - Thomas Bodner - Wolfgang Lehner - Scott MacLean - Peter Bumbulis

Real-Time Analytical Processing with SQL Server

Paul Larson - Adrian Birka - Eric Hanson - Weiyun Huang - Michal Novakiewicz - Vassilis Papadimos

The Dataflow Model: A Practical Approach to Balancing Correctness, Latency, and Cost in Massive-Scale, Unbounded, Out-of-Order Data Processing

Tyler Akidau - Robert Bradshaw - Craig Chambers - Slava Chernyak - Rafael Fernandez-Moctezuma - Reuven Lax - Sam McVeety - Daniel Mills - Frances Perry - Eric Schmidt - Sam Whittle

Live Programming Support in the LogicBlox System

Todd Green - Dan Olteanu - Geoffrey Washburn

Indexing and Selecting Hierarchical Business Logic

Anja Gruenheid - Alessandra Loro - Donald Kossman - Damien Profeta - Philippe Beaudequin

Distributed Architecture of Oracle Database In-memory

Niloy Mukherjee - Shasank Chavan - Maria Colgan - Dinesh Das - Mike Gleeson - Sanket Hase - Allison Holloway - Hui Jin - Jesse Kamp - Kartik Kulkarni - Tirthankar Lahiri - Juan Loaiza - Vineet Marwah - Andy Witkowski - Jiaqi Yan - Mohamed Zait

Gorilla: Facebook's Fast, Scalable, In-Memory Time Series Database

Justin Teller - Scott Franklin - Tuomas Pelkonen - Paul Cavallaro

Query Optimization in Oracle 12c Database In-Memory

Dinesh Das - Jiaqi Yan - Mohamed Zait - Satya Valluri - Nirav Vyas - Ramarajan Krishnamachari - Prashant Gaharwar - Jesse Kamp - Niloy Mukherjee

Building a Replicated Logging System with Apache Kafka

Guozhang Wang - Joel Koshy - Sriram Subramanian - Kartik Paramasivam - Mammad Zadeh - Neha Narkhede - Jun Rao - Jay Kreps - Joe Stein

Optimization of Common Table Expressions in MPP Database Systems

Amr El-Helw - Venkatesh Raghavan - Mohamed Soliman - George Caragea - Zhongxian Gu - Michalis Petropoulos

One Trillion Edges: Graph Processing at Facebook-Scale

Avery Ching - Dionysios Logothetis - Sergey Edunov - Maja Kabiljo - Sambavi Muthukrishnan

Differential Privacy in Telco Big Data Platform

Xueyang Hu - Mingxuan Yuan - Jianguo Yao - Yu Deng - Lei Chen - Haibing Guan - Jia Zeng

Efficient Evaluation of Object-Centric Exploration Queries for Visualization

You Wu - Boulos Harb - Jun Yang - Cong Yu

ACME: A Parallel Cloud-Oriented System for Extracting Frequent Patterns from a Very Long Sequence

Majed Sahl

Efficient Distributed Subgraph Similarity Matching

Ye Yuan

Efficient k-Closest Pair Queries in General Metric Spaces

Yunjun Gao

Task-Assignment Optimization in Knowledge Intensive Crowdsourcing

Senjuti Basu Roy

Data Profiling - A Survey

Felix Naumann

Data Generation for Testing and Grading SQL Queries

Bikash Chandra/S Sudarshan

Wednesday Sep 2nd 07:30-08:30

Light Breakfast

Location: Grand Promenade

Wednesday Sep 2nd 08:30-10:00

Industrial Keynote: Todd Walter; Academic Keynote: Magdalena Balazinska

Location: Monarchy Ballroom

Chair: Chen Li

Big Plateaus of Big Data on the Big Island

Todd Walter, Teradata



Bio: Todd Walter is the Chief Technologist for Teradata across the Americas region. With substantive expertise in big data, database engineering and systems architecture, he works closely with Teradata customers, colleagues, and alliance partners to evaluate and prioritize initiatives — and implement data strategy and analytics. As a pragmatic visionary, Walter helps customer business analysts as well as technologists better understand all of the astonishing possibilities of big data and analytics in view of emerging as well as existing capabilities of information infrastructures. Todd works with organizations of all sizes and levels of experience, from start-ups to Fortune 100 companies at the leading edge of adopting big data, data warehouse and analytics technologies. Walter has been with Teradata for nearly 28 years, contributing significantly to Teradata's unique design features and functionality. He holds more than a dozen Teradata patents and is a Teradata Fellow, the highest technical award granted by the company. Todd served for more than ten years as Chief Technical Officer of Teradata Labs, responsible for vision, strategy and technical leadership of the Teradata product line before taking on his current strategic consulting role.

Big Data Research: Will Industry Solve all the Problems?

Magdalena Balazinska, University of Washington



Bio: Magdalena Balazinska is an Associate Professor in the department of Computer Science and Engineering at the University of Washington and the Jean Loup Baer Professor of Computer Science and Engineering. She's the director of the IGERT PhD Program in Big Data and Data Science. She's also a Senior Data Science Fellow of the University of Washington eScience Institute. Magdalena's research interests are in the field of database management systems. Her current research focuses on big data management, scientific data management, and cloud computing. Magdalena holds a Ph.D. from the Massachusetts Institute of Technology (2006). She is a Microsoft Research New Faculty Fellow (2007), received an NSF CAREER Award (2009), a 10-year most influential paper award (2010), an HP Labs Research Innovation Award (2009 and 2010), a Rogel Faculty Support Award (2006), a Microsoft Research Graduate Fellowship (2003-2005), and multiple best-paper awards.

Wednesday Sep 2nd 10:00-10:30

Refreshments: Coffee Break

Location: Grand Promenade

Wednesday Sep 2nd 10:30-12:15

Research 9: Graph Processing 2

Location: Kings 1

Chair: Peter Triantafyllou

Scalable Subgraph Enumeration in MapReduce

Longbin Lai (UNSW), Lu Qin (University of Technology (Sydney)), Xuemin Lin (University of New South Wales), Lijun Chang (University of New South Wales)

FrogWild! -- Fast PageRank Approximations on Graph Engines

Ioannis Mitliagkas (UT Austin), Michael Borokhovich (UT Austin), Alexandros Dimakis (UT Austin), Constantine Caramanis (UT Austin)

Pregel Algorithms for Graph Connectivity Problems with Performance Guarantees

Da Yan (HKUST), James Cheng (CUHK), Kai Xing (HKUST), Yi Lu (CUHK), Wilfred Ng, Yingyi Bu (UC Irvine)

LogGP: A Log-based Dynamic Graph Partitioning Method

Ning Xu, Lei Chen (Hong Kong University of Science and Technology), Bin Cui (Peking University)

Blogel: A Block-Centric Framework for Distributed Computation on Real-World Graphs

Da Yan (HKUST), James Cheng (CUHK), Yi Lu (CUHK), Wilfred Ng, The Hong Kong University of Science and Technology)

Research 10: Novel DB Architectures

Location: Kings 2

Chair: Sharad Mehrotra

Coordination autoance in Database Systems

Peter Bailis (UC Berkeley), Alan Fekete (University of Sydney), Michael Franklin (UC Berkeley), Ali Ghodsi (UC Berkeley), Joseph Hellerstein (UC Berkeley), Ion Stoica (UC Berkeley)

A Scalable Search Engine for Mass Storage Smart Objects

Nicolas Ancaux (INRA and University of Versailles Saint-Quentin), Saliha Lallali (INRIA and University of Versailles Saint-Quentin), Iulian Sandu Popa (University of Versailles), Philippe Pucheral (INRIA and University of Versailles Saint-Quentin)

Principles of Dataset Versioning: Exploring the Recreation/Storage Tradeoff

Souvik Bhattacharjee (University of Maryland (College Park)), Amit Chavan (University of Maryland at College Park), Silu Huang (University of Illinois at Urbana-Champaign), Amol Deshpande (University of Maryland at College Park), Aditya Parameswaran, University of Illinois at Urbana-Champaign)

Schema Management for Document Stores

Lanjuan Wang (IBM Research-China), Oktie Hassanzadeh (IBM Research), Shuo Zhang (IBM Research-China), Juwei Shi (IBM Research-China), Limei Jiao, Jia Zou (IBM Research-China), Chen Wang (Tsinghua University)

Supporting Scalable Analytics with Latency Constraints

Boduo Li (University of Massachusetts Amherst), Yanlei Diao (University of Massachusetts Amherst), Prashant Shenoy (University of Massachusetts Amherst)

Industrial 3: Real-time and Interactive Analytics

Chair: TBA

JetScope: Reliable and Interactive Analytics at Cloud Scale

Eric Boutin (Microsoft), Paul Brett (Microsoft), Xiaoyu Chen (Microsoft), Jaliya Ekanayake (Microsoft), Tao Guan (Microsoft), Anna Korsun (Microsoft), Zhicheng Yin (Microsoft), Nan Zhang (Microsoft), Jingren Zhou (Microsoft)

Towards Scalable Real-time Analytics: An Architecture for Scale-out of OLXP Workloads

Anil Goel (SAP Labs), Jeffrey Pound (SAP Labs), Nathan Auch (SAP Labs), Peter Bumbulis (SAP Labs), Scott MacLean (SAP Labs), Franz Faerber (SAP SE), Francis Gropengiesser (SAP SE), Christian Mathis (SAP SE), Thomas Bodner (SAP SE), Wolfgang Lehner (TU Dresden)

Real-Time Analytical Processing with SQL Server

Paul Larson (Microsoft), Adrian Birka (Microsoft), Eric Hanson (Microsoft), Weiyun Huang (Microsoft), Michal Novakiewicz (Microsoft), Vassilis Papadimos (Microsoft)

Research 11: Social Network Analysis

Location: Queens 4

Chair: Laks Lakshmanan

Inferring Continuous Dynamic Social Influence and Personal Preference for Temporal Behavior Prediction

Jun Zhang (Tsinghua University), Chaokun Wang (Tsinghua University), Jianmin Wang (Tsinghua University), Jeffrey Xu Yu (The Chinese University of Hong Kong (Hong Kong))

Influential Community Search in Large Networks

Rong-Hua LI (CUHK), Lu Qin (University of Technology (Sydney)), Jeffrey Xu Yu (The Chinese University of Hong Kong), Rui Mao (Shenzhen University)

Linearized and Single-Pass Belief Propagation

Wolfgang Gatterbauer (Carnegie Mellon University), Stephan Günnemann (Carnegie Mellon University), Danai Koutra (Carnegie Mellon University), Christos Faloutsos (Carnegie Mellon University)

Online Topic-Aware Influence Maximization

Shuo Chen (Tsinghua University), Ju Fan (National University of Singapore), Guoliang Li (Tsinghua University), Jianhua Feng (Tsinghua University), Kian-Lee Tan (National University of Singapore), Jinhui Tang (National University of Singapore)

Walk, Not Wait: Faster Sampling Over Online Social Networks

Azade Nazi (University of Texas at Arlington), Zhuojie Zhou (George Washington University), Saravanan Thirumuruganathan (University of Texas at Arlington), Nan Zhang (George Washington University), Gautam Das (University of Texas at Arlington)

Research 12: Query Processing 1

Location: Queens 5

Chair: Dan Olteanu

Work-Efficient Parallel Skyline Computation for the GPU

Kenneth Bøgh (Århus Universitet), Sean Chester (Århus Universitet), Ira Assent (Århus Universitet)

Memory-Efficient Hash Joins

Gopi Attaluri (IBM Software Group), Ronald Barber (IBM Research-Almaden), Naresh Chainani (IBM Software Group), Sam Lightstone (IBM Software Group), Guy Lohman (IBM Research-Almaden), Ippokratis Pandis (IBM Research-Almaden), Vijayshankar Raman (IBM Research-Almaden), Dave Sharpe (IBM Software Group), Richard Sidle (IBM Research-Almaden)

MRCSI: Compressing and Searching String Collections with Multiple References

Sebastian Wandelt (HU Berlin), Ulf Leser (HU Berlin)

Trill: A High-Performance Incremental Query Processor for Diverse Analytics

Badrish Chandramouli (Microsoft Research), Jonathan Goldstein (Microsoft Research), Mike Barnett (Microsoft Research), Robert DeLine (Microsoft Research), Danyel Fisher (Microsoft Research), John Platt (Microsoft Research), James Terwilliger (Microsoft Research), John Wernsing (Microsoft Research)

Rapid Sampling for Visualizations with Ordering Guarantees

Albert Kim (MIT), Eric Blais (MIT), Aditya Parameswaran (MIT and U Illinois (UIUC)), Piotr Indyk (MIT), Sam Madden (MIT), Ronitt Rubinfeld (MIT and Tel Aviv University)

Tutorial 3: Structured Analytics in Social Media

Location: Queens 6

Structured Analytics in Social Media

Mahashweta Das, Gautam Das

Demo 3: Systems, User Interfaces, and Visualization

Location: Kona 4

Chair: Tianli Mo

FP-Hadoop: Efficient Execution of Parallel Jobs Over Skewed Data

Miguel Liroz-Gistau (INRIA), Reza Akbarinia (INRIA), Patrick Valduriez (INRIA)

SDB: A Secure Query Processing System with Data Interoperability

Zhian He (Hong Kong Polytechnic University), WaiKit Wong (Hang Seng Management College), Ben Kao (University of Hong Kong), David W. Cheung (University of Hong Kong), Rongbin Li (University of Hong Kong), Siu Ming Yiu (University of Hong Kong), Eric Lo (Polytechnic University of Hong Kong)

A Demonstration of HadoopViz: An Extensible MapReduce-based System for Visualizing Big Spatial Data

Ahmed Eldawy (University of Minnesota), Mohamed Mokbel (University of Minnesota), Christopher Jonathan (University of Minnesota)

A Demonstration of the BigDAWG Multi-Database System

Aaron Elmore (MIT), Jennie Duggan (Northwestern), Michael Stonebraker (MIT), Manasi Vartak (MIT), Sam Madden (MIT), Vijay Gadepally (MIT), Jeremy Kepner (MIT), Timothy Mattson (Intel), Jeff Parhurst (Intel), Stavros Papadopoulos (MIT), Nesime Tatbul (Intel Labs and MIT), Magdalena Balazinska (University of Washington), Bill Howe (University of Washington), Jeffrey Heer (University of Washington), David Maier (Portland State University), Tim Kraska (Brown), Ugur Cetintemel (Brown University), Stan Zdonik (Brown University)

RINSE: Interactive Data Series Exploration

Kostas Zoumpatianos (University of Trento), Stratos Idreos (Harvard), Themis Palpanas (Paris Descartes University)

Smart Drill-Down: A New Data Exploration Operator

Manas Joglekar (Stanford University), Hector Garcia-Molina (Stanford University), Aditya Parameswaran (University of Illinois at Urbana Champaign)

VIQ: auto-suggestion enabled visual interface for interactive graph query formulation

Nandish Jayaram (University of Texas at Arlington), Sidharth Goyal (University of Texas at Arlington), Chengkai Li (University of Texas at Arlington)

VINERY: A Visual IDE for Information Extraction

Yunyao Li (IBM Research-Almaden), Elmer Kim (Treasurer Data (Inc.)), Marc Touchette (IBM Silicon Valley Lab), Ramiya Venkatachalam (IBM Silicon Valley Lab), Hao Wang (IBM Silicon Valley Lab)

GIS navigation boosted by column stores

Foteini Alvanaki (CWI), Romulo Goncalves (Netherlands eScience Center), Milena Ivanova (NuoDB), Martin Kersten (CWI), Kostis Kyzirakos (CWI)

AIDE: An Automatic User Navigation System for Interactive Data Exploration

Yanlei Diao (University of Massachusetts Amherst), Kyriaki Dimitriadou (Brandeis University), Zhan Li (Brandeis University), Wenzhao Liu (University of Massachusetts Amherst), Olga Papaemmanouil (Brandeis University), Kemi Peng (Brandeis University), Liping Peng (University of Massachusetts Amherst)

A Demonstration of AQWA: Adaptive Query-Workload-Aware Partitioning of Big Spatial Data

Ahmed Aly (Purdue University), Ahmed Abdelhamid (Purdue University), Ahmed Mahmood (Purdue University), Walid Aref (Purdue University), Mohamed Hassan (Purdue University), Hazem Elmeleegy (Turn Inc), Mourad Ouzzani (Qatar Computing Research Institute)

Data-Spread: Unifying Databases and Spreadsheets

Mangesh Bendre (University of Illinois at Urbana-Champaign), Bofan Sun (University of Illinois at Urbana-Champaign), Ding Zhang (University of Illinois at Urbana-Champaign), Xinyan Zhou (University of Illinois at Urbana-Champaign), Kevin Chang (University of Illinois at Urbana-Champaign), Aditya Parameswaran (University of Illinois at Urbana-Champaign)

CODD: A Dataless Approach to Big Data Testing

Ashoke S (Indian Institute of Science), Jayant Haritsa (IISc)

Vizdom: Interactive Analytics through Pen and Touch

Andrew Crotty (Brown University), Alex Galakatos (Brown University), Emanuel Zraggen (Brown University), Carsten Binnig (Brown University), Tim Kraska (Brown University)

DBSeer: Pain-free Database Administration through Workload Intelligence

Dong Young Yoon (University of Michigan Ann Arbor), Barzan Mozafari (University of Michigan Ann Arbor), Douglas Brown (Teradata Inc.)

Wednesday Sep 2nd 12:15-13:30

Lunch

Location: Lagoon Lanai

Wednesday Sep 2nd 13:30-15:00

Research 13: Graph Processing 3

Location: Kings 1

Chair: Mohamed Mokbel

TOP: A Framework for Enabling Algorithmic Optimizations for Distance-Related Problems

Yufei Ding (North Carolina State University), Xipeng Shen (North Carolina State University), Madanlal Musuvathi (Microsoft Research), Todd Mytkowicz (Microsoft Research)

SCAN++: Efficient Algorithm for Finding Clusters, Hubs and Outliers on Large-scale Graphs

Hiroaki Shiohara (NTT), Yasuhiro Fujiwara (NTT), Makoto Onizuka (Osaka University)

GraphTwist: Fast Iterative Graph Computation with Two-tier Optimizations

Yang Zhou (Georgia Institute of Technology), Ling Liu (Georgia Institute of Technology), Kisung Lee (Georgia Institute of Technology), Qi Zhang (Georgia Institute of Technology)

A Scalable Distributed Graph Partitioner

Daniel Margo (Harvard University), Margo Seltzer (Harvard University)

Keys for Graphs

Wenfei Fan (University of Edinburgh), Zhe Fan (University of Edinburgh), Chao Tian (University of Edinburgh), Xin Luna Dong (Google Inc)

Research 14: Novel Hardware Architectures

Location: Kings 2

Chair: Paul Larson

Scaling Up Concurrent Main-Memory Column-Store Scans: Towards Adaptive NUMA-aware Data and Task Placement

Iraklis Psaroudakis (EPFL), Tobias Scheuer (SAP SE), Norman May (SAP AG), Abdelkader Sellami (SAP SE), Anastassia Ailamaki (EPFL)

In-Memory Performance for Big Data

Goetz Graefe (HP Labs), Haris Volos (HP Labs), Hideaki Kimura (HP Labs), Harumi Kuno (HP Labs), Joseph Tucek (HP Labs), Mark Lillibridge (HP Labs), Alistair Veitch (Google)

Profiling R on a Contemporary Processor

Shriram Sridharan (University of Wisconsin-Madison), Jignesh Patel (University of Wisconsin-Madison)

Deployment of Query Plans on Multicores

Jana Giceva (ETH Zurich), Gustavo Alonso (ETH Zurich), Timothy Roscoe (ETH Zurich), Tim Harris (Oracle labs)

Faster Set Intersection with SIMD instructions by Reducing Branch Mispredictions

Hiroshi Inoue (IBM Research-Tokyo and University of Tokyo), Moriyoshi Ohara (IBM Research-Tokyo), Kenjiro Taura (University of Tokyo)

Industrial 4: Novel Approaches to Modern Data Processing

Location: Kings 3

Chair: John Cieslewicz

The Dataflow Model: A Practical Approach to Balancing Correctness, Latency, and Cost in Massive-Scale, Unbounded, Out-of-Order Data Processing

Tyler Akidau (Google), Robert Bradshaw (Google), Craig Chambers (Google), Slava Chernyak (Google), Rafael Fernández-Moctezuma (Google), Reuven Lax (Google), Sam McVeety (Google), Daniel Mills (Google), Frances Perry (Google), Eric Schmidt (Google), Sam Whittle (Google)

Live Programming Support in the LogicBlox System

Todd Green (LogicBlox Inc.), Dan Olteanu (LogicBlox Inc.), Geoffrey Washburn (LogicBlox Inc.)

Indexing and Selecting Hierarchical Business Logic

Alessandra Loro (Palantir Technologies), Anja Gruenheid (ETH Zurich), Donald Kossman (ETH Zurich and Microsoft Research), Damien Profeta (S.A.S. Amadeus), Philippe Beaudequin (S.A.S. Amadeus)

Research 15: Query Optimization

Location: Queens 4

Chair: Jonathan Goldstein

Resource Bricolage for Parallel Database Systems

Jiexing Li (Google Inc), Jeffrey Naughton (University of Wisconsin-Madison), Rimma Nehme (Microsoft Jim Gray Systems Lab)

Multi-Objective Parametric Query Optimization

Immanuel Trummer (EPFL), Christoph Koch (EPFL)

Querying with Access Patterns and Integrity Constraints

Michael Benedikt (Oxford University), Julien Leblay (Oxford University), Efi Tsamoura (Oxford University)

Uncertainty Aware Query Execution Time Prediction

Wentao Wu (University of Wisconsin-Madison), Xi Wu (University of Wisconsin-Madison), Hakan Hacigumus (NEC Labs America), Jeffrey Naughton (University of Wisconsin-Madison)

Join Size Estimation Subject to Filter Conditions

David Vengerov (Oracle Labs), Andre Menck (Oracle Corp.), Mohamed Zait (Oracle Corp), Sunil Chakkappen (Oracle Corp)

Research 16: Crowdsourcing and Social Network Analysis

Location: Queens 5

Chair: Yi Chen

Scaling Up Crowd-Sourcing to Very Large Datasets: A Case for Active Learning

Barzan Mozafari (University of Michigan), Purna Sarkar (UC Berkeley), Michael Franklin (UC Berkeley), Michael Jordan (UC Berkeley), Sam Madden (MIT)

Hear the Whole Story: Towards the Diversity of Opinion in Crowdsourcing Markets

Ting Wu (Hong Kong University of Science and Technology), Lei Chen (Hong Kong University of Science and Technology), Pan Hui (Hong Kong University of Science and Technology), CHEN ZHANG (Hong Kong University of Science and Technology), Weikai Li (Hong Kong University of Science and Technology)

Where To: Crowd-Aided Path Selection

Chen Zhang (Hong Kong University of Science and Technology), Yongxin Tong (Hong Kong University of Science and Technology), Lei Chen (Hong Kong University of Science and Technology)

Reliable Diversity-Based Spatial Crowdsourcing by Moving Workers

Peng Cheng (Hong Kong University of Science and Technology), Xiang Lian (University of Texas Rio Grande Valley), Zhao Chen (Hong Kong University of Science and Technology), Rui Fu (Hong Kong University of Science and Technology), Lei Chen (Hong Kong University of Science and Technology), Jinsong Han (Xi'an Jiaotong University), Jizhong Zhao (Xi'an Jiaotong University)

Learning User Preferences By Adaptive Pairwise Comparison

Li Qian (Facebook), Jinyang Gao (National University of Singapo), H V Jagadish (University of Michigan Ann Arbor)

Tutorial 4: SQL-on-Hadoop Systems (1/2)

Location: Queens 6

SQL-on-Hadoop Systems (1/2)

Daniel Abadi, Shivnath Babu, Fatma Ozcan, Ippokratis Pandis

Panel 1: 40-years VLDB

Location: Kona 1-2-3

Chair: Pat Selinger

40-years VLDB

Phil Bernstein (Microsoft Research), Michael Brodie (MIT and retired Chief Scientist Verizon IT), Don Chamberlin (retired IBM Fellow), Alfons Kemper (Technical University Munich), Michael Stonebraker (MIT and serial entrepreneur), Pat Selinger (Paradata)



Bio: Philip A. Bernstein is a Distinguished Scientist at Microsoft Research. Over the past 35 years, he has been a product architect at Microsoft and Digital Equipment Corp., a professor at Harvard University and Wang Institute of Graduate Studies, and a VP Software at Sequoia Systems. He has published over 150 papers and two books on the theory and implementation of database systems, especially on transaction processing and metadata management. His latest work focuses on database systems and object-oriented middleware for distributed computing, and integration of heterogeneous data in the enterprise and on the web. He is an ACM Fellow, a winner of the ACM SIGMOD Innovations Award, and a member of the National Academy of Engineering. He received a B.S. from Cornell and M.Sc. and Ph.D. from University of Toronto. His home page is: <http://research.microsoft.com/~philbe>



Bio: Dr. Brodie has over 40 years experience in research and industrial practice in databases, distributed systems, integration, artificial intelligence, and multi-disciplinary problem solving. He is concerned with the Big Picture aspects of information ecosystems including business, economic, social, application, and technical. Dr. Brodie is a Research Scientist, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology; advises startups; serves on Advisory Boards of national and international research organizations; and is an adjunct professor at the National University of Ireland, Galway and at the University of Technology, Sydney. For over 20 years he served as Chief Scientist of IT, Verizon, a Fortune 15 company, responsible for advanced technologies, architectures, and methodologies for IT strategies and for guiding industrial scale deployments of emerging technologies. His current research and applied interests include Big Data, Data Science, data curation at scale and a related start up Tamr.com. He has served on several National Academy of Science committees. Dr. Brodie holds a PhD in Databases from the University of Toronto and a Doctor of Science (honoris causa) from the National University of Ireland.



Bio: Don Chamberlin is co-inventor, with Ray Boyce, of SQL, the world's most widely-used database query language. He was also one of the managers of System R, the IBM research project that produced the first implementation of SQL. More recently, Don represented IBM on the W3C working group that developed XQuery, a query language for XML data. Don received his B.S. degree from Harvey Mudd College and his Ph.D. from Stanford University. He has been named a Fellow of IBM, ACM, IEEE, and the Computer History Museum, and has received the ACM Software Systems Award and the SIGMOD Innovations Award. For several years Don has contributed problems to the annual ACM International Collegiate Programming Contest. He has also served as an adjunct professor of computer science at University of California, Santa Cruz, and at Santa Clara University. Don is currently retired and is dividing his time among learning, traveling, volunteer activities, and enjoying his grandchildren.



Bio: Alfons Kemper's research field is database systems engineering. He explores ways to optimize information systems for operational and scientific applications as a way to combat the data explosion. His main areas of interest are optimization concepts for distributed information structures, data integration methods and, in particular, main memory-based database systems. Together with his colleague Thomas Neumann he leads the HyPer main-memory database project (hyper-db.com) at Technische Universität München. HyPer is one of the first hybrid database systems that offers high-performance OLTP as well as OLAP in parallel on the same database state. After studying computer science at the University of Dortmund from 1977 to 1980, he moved to the University of Southern California, Los Angeles. While there, he obtained his Master of Science and doctorate. Upon his return to Germany, he completed his lecturer qualification at the University of Karlsruhe. His first professorship was conferred by RWTH Aachen. After many years as Director of the Chair of Database Systems at the University of Passau, TUM offered him a position in 2004. From 2006 to 2010, he was Dean of the Department of Informatics at TUM. His textbook on database systems, published by deGruyter and now in its 10th edition, is a best-seller in German-speaking countries and is used in most universities and colleges.



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



Bio: Dr. Pat Selinger is the Chief Technology Officer at Paradata (Paradata.io) where she is working on challenging problems in data harmonization, curation, provenance, and entity resolution. Prior to joining Paradata, she worked at IBM Research. She is a world-renowned pioneer in relational database management and inventor of the technique of cost-based query. She was a key member of the original System R team that created the first relational database research prototype. She also established and led IBM's Database Technology Institute, considered one of the most successful examples of a fast technology pipeline from research to development and personally has technical contributions in the areas of database optimization, data parallelism, distributed data, and unstructured data management. Dr. Selinger was appointed an IBM Fellow in 1994 and is an ACM Fellow, a member of the National Academy of Engineering, and a Fellow of the American Academy of Arts and Sciences. Dr. Selinger has also received the ACM Systems Software Award for her work on System R and has received the SIGMOD Innovation Award.

Wednesday Sep 2nd 15:00-15:30

Refreshments: Coffee Break
Location: Grand Promenade

Wednesday Sep 2nd 15:30-17:00

Research 17: Graph Processing Systems
Location: Kings 1
Chair: Tilmann Rabl

Pregelix: Big(ger) Graph Analytics on A Dataflow Engine

Yingyi Bu (UC Irvine), Vinayak Borkar (X15 Software Inc), Jianfeng Jia (UC Irvine), Michael Carey (UC Irvine), Tyson Condie (UCLA)

Large-Scale Distributed Graph Computing Systems: An Experimental Evaluation

Yi Lu (CUHK), James Cheng (CUHK), Da Yan (HKUST), Huanhuan Wu (CUHK)

Fast Failure Recovery in Distributed Graph Processing Systems

Yanyan Shen (National University of Singapore), Gang Chen (Zhejiang University), H V Jagadish (University of Michigan Ann Arbor), Wei Lu (Renmin University), Beng Chin Ooi (National University of Singapore), Bogdan Tudor (National University of Singapore)

Giraph Unchained: Barrierless Asynchronous Parallel Execution in Pregel-like Graph Processing Systems

Minyang Han (University of Waterloo), Khuzaima Daudjee (University of Waterloo)

GraphMat: High performance graph analytics made productive

Narayanan Sundaram (Intel Labs), Nadathur Satish (Intel Labs), Mostofa Ali Patwary (Intel Labs), Subramanya Dulloor (Intel Labs), Michael Anderson (Intel Labs), Satya Gautam Vadlamudi (Intel Labs), Dipankar Das (Intel Labs), Pradeep Dubey (Intel Labs)

Research 18: Novel Hardware Architectures 2

Location: Kings 2
Chair: Jayant Haritsa

In-Cache Query Co-Processing on Coupled CPU-GPU Architectures

Jiong He (NTU), Shuhao Zhang (NTU), Bingsheng He (NTU)

NVRAM-aware Logging in Transaction Systems

Jian Huang (Georgia Tech), Karsten Schwan (Georgia Tech), Moinuddin Qureshi (Georgia Tech)

Improving Main Memory Hash Joins on Intel Xeon Phi Processors: An Experimental Approach

Saurabh Jha (Nanyang Technological University), Bingsheng He (Nanyang Technological University), Mian Lu (A*STAR IHPC), Xuntao Cheng (Nanyang Technological University), Phung Huynh Huynh (A*STAR IHPC)

REWIND: Recovery Write-Ahead System for In-Memory Non-Volatile Data-Structures

Andreas Chatzistergiou (University of Edinburgh), Marcelo Cintra (Intel), Stratis Vlgas (University of Edinburgh)

Persistent B+-Trees in Non-Volatile Main Memory

Shimin Chen (Chinese Academy of Sciences), Qin Jin (Renmin University)

Industrial 5: In-memory Data Management

Location: Kings 3

Chair: Amr El-Helw

Distributed Architecture of Oracle Database In-memory

Niloy Mukherjee (Oracle Corporation), Shasank Chavan (Oracle Corporation), Maria Colgan (Oracle Corporation), Dinesh Das (Oracle Corporation), Mike Gleeson (Oracle Corporation), Sanket Hase (Oracle Corporation), Allison Holloway (Oracle Corporation), Hui Jin (Oracle Corporation), Jesse Kamp (Oracle Corporation), Kartk Kulkarni (Oracle Corporation), Tirthankar Lahiri (Oracle Corporation), Juan Loaiza (Oracle Corporation), Vineet Marwah (Oracle Corporation), Andy Witkowski (Oracle Corporation), Jiaqi Yan (Oracle Corporation), Mohamed Zait (Oracle Corporation)

Gorilla: Facebook's Fast, Scalable, In-Memory Time Series Database

Justin Teller (Facebook), Scott Franklin (Facebook), Tuomas Pelkonen (Facebook), Paul Cavallaro (Facebook)

Query Optimization in Oracle 12c Database In-Memory

Dinesh Das (Oracle), Jiaqi Yan (Oracle), Mohamed Zait (Oracle Corp), Satyanarayana Valluri, EPFL, Nirav Vyas (Oracle), Ramarajan Krishnamachari (Oracle), Prashant Gaharwar (Oracle), Jesse Kamp (Oracle), Niloy Mukherjee (Oracle Corporation)

Research 19: Social Network Analysis

Location: Queens 4

Chair: Felix Naumann

Robust Local Community Detection: On Free Rider Effect and Its Elimination

Yubao Wu (Case Western Reserve University), Ruoming Jin (Kent State University), Jing Li (Case Western Reserve University), Xiang Zhang (Case Western Reserve University)

Viral Marketing Meets Social Advertising: Ad Allocation with Minimum Regret

Cigdem Aslay (University Pompeu Fabra), Wei Lu (University of British Columbia), Francesco Bonchi (Yahoo Labs), Amit Goyal (Twitter), Laks Lakshmanan (University of British Columbia)

Community Detection in Social Networks: An In-depth Benchmarking Study with a Procedure-Oriented Framework

Meng Wang (Tsinghua University), Chaokun Wang (Tsinghua University), Jeffrey Xu Yu (The Chinese University of Hong Kong), Jun Zhang (Tsinghua University)

Leveraging History for Faster Sampling of Online Social Networks

Zhuojie Zhou (George Washington University), Nan Zhang (George Washington University), Gautam Das (University of Texas at Arlington)

Real-time Targeted Influence Maximization for Online Advertisements

Yuchen Li (National University of Singapore), Dongxiang Zhang (National University of Singapore), Kian-Lee Tan (National University of Singapore)

Research 20: Ranking and Top-K

Location: Queens 5

Chair: Raymond Wong

Top-k Nearest Neighbor Search In Uncertain Data Series

Michele Dallachiesa (University of Trento), Themis Palpanas (Paris Descartes University), Ihab Ilyas (QCRI)

Scaling Manifold Ranking Based Image Retrieval

Yasuhiro Fujiwara (NTT), Go Irie (NTT), Shari Kuroyama (California Institute of Technology), Makoto Onizuka (Osaka University)

Optimal Enumeration: Efficient Top-k Tree Matching

Lijun Chang (University of New South Wales), Xuemin Lin (University of New South Wales), Wenjie Zhang (UNSW), Jeffrey Xu Yu (The Chinese University of Hong Kong (Hong Kong)), Ying Zhang (University of Technology (Sydney)), Lu Qin (University of Technology (Sydney))

Generating Top-k Packages via Preference Elicitation

Min Xie (University of British Columbia), Laks V. S. Lakshmanan (University of British Columbia), Peter Wood (Birkbeck (University of London))

Rank aggregation with ties: Experiments and Analysis

Bryan Brancotte (Université Paris Sud), Bo Yang (Wuhan University (Wu)), Guillaume Blin (University of Bordeaux), Sarah Cohen-Boulakia (Université Paris Sud), Sylvie Hamel (Université de Montréal)

Tutorial 5: SQL-on-Hadoop Systems (2/2)

Location: Queens 6

SQL-on-Hadoop Systems (2/2)

Daniel Abadi, Shivnath Babu, Fatma Ozcan, Ippokratis Pandis

Demo 1: Data Mining, Graph, Text, and Semi-structured Data

Location: Kona 4

Chair: Tianli Mo

Evaluating SPARQL Queries on Massive RDF Datasets

Razen Harbi (King Abdullah University of Science and Technology), Ibrahim Abdelaziz (King Abdullah University of Science and Technology), Panos Kalnis (King Abdullah University of Science and Technology), Nikos Mamoulis (University of Ioannina)

Demonstration of Santoku: Optimizing Machine Learning over Normalized Data

Arun Kumar (University of Wisconsin-Madison), Mona Jalal (University of Wisconsin-Madison), Boqun Yan (University of Wisconsin-Madison), Jeffrey Naughton (University Wisconsin-Madison), Jignesh Patel (University of Wisconsin-Madison)

PRISM: Concept-preserving Summarization of Top-K Social Image Search Results

Boon-Siew Seah (Nanyang Technological University), Sourav S Bhowmick (Nanyang Technological University), Aixin Sun (Nanyang Technological University)

SPARTEX: A Vertex-Centric Framework for RDF Data Analytics

Ibrahim Abdelaziz (King Abdullah University of Science and Technology), Razen Harbi (King Abdullah University of Science and Technology), Semih Salihoglu (Stanford University), Panos Kalnis (King Abdullah University of Science and Technology), Nikos Mamoulis (University of Ioannina)

I2RS: A Distributed Geo-Textual Image Retrieval and Recommendation System

Lu Chen (Zhejiang University), Yunjun Gao (Zhejiang University), Zhihao Xing (Zhejiang University), Christian Jensen (Aalborg University), Gang Chen (Zhejiang University)

Reformulation-based query answering in RDF: alternatives and performance

Damian Bursztyn (INRIA), Francois Goasdoue (University of Rennes 1), Ioana Manolescu (INRIA)

TreeScope: Finding Structural Anomalies In Semi-Structured Data

Shanshan Ying (ADSC), Flip Korn, Barna Saha (University of Massachusetts Amherst), Divesh Srivastava (AT&T Labs-Research)

PERSEUS: An Interactive Large-Scale Graph Mining and Visualization Tool

Danai Koutra (Carnegie Mellon University), Di Jin (Carnegie Mellon University), Yuanchi Ning (Uber Technologies Inc.), Christos Faloutsos (Carnegie Mellon University)

Virtual eXist-db: Liberating Hierarchical Queries from the Shackles of Access Path Dependence

Curtis Dyreson (Utah State University), Sourav S Bhowmick (Nanyang Technological University), Ryan Grapp (Utah State University)

FLORIN - A System to Support (Near) Real-Time Applications on User Generated Content on Daily News

Qingyuan Liu (Temple University), Eduard Dragut (Temple University), Arjun Mukherjee (University of Houston), Weiyi Meng (Binghamton University)

A Framework for Clustering Uncertain Data

Erich Schubert (Ludwig-Maximilians-Universität München), Alexander Koos (Ludwig-Maximilians-Universität München), Tobias Emrich (Ludwig-Maximilians-Universität München), Andreas Züfle (Ludwig-Maximilians-Universität München), Klaus Schmid (Ludwig-Maximilians-Universität München), Arthur Zimek (Ludwig-Maximilians-Universität München)

Query-oriented summarization of RDF graphs

Sejla Cebiric (INRIA), Francois Goasdoue (University of Rennes 1), Ioana Manolescu (INRIA)

Universal-DB: Towards Representation Independent Graph Analytics

Yodsawalai Chodpathumwan (University of Illinois), Amirhossein Aleyassin (University of Illinois), Arash Termehchy (Oregon State University), Yizhou Sun (Northeastern University)

Tornado: A Distributed Spatio-Textual Stream Processing System

Ahmed Mahmood (Purdue University), Ahmed Aly (Purdue University), Thamir Qadah (Purdue University), El Kindi Rezig (Purdue University), Anas Daghistani (Purdue University), Amgad Madkour (Purdue University), Ahmed Abdelhamid (Purdue University), Mohamed Hassan (Purdue University), Walid Aref (Purdue University (USA)), Saleh Basalamah (Umm Al-Qura University)

S+EPP: Construct and Explore Bisimulation Summaries, plus Optimize Navigational Queries; all on Existing SPARQL Systems

Mariano Consens (University of Toronto), Valeria Fionda (University of Calabria), Shahan Khatchadourian (University of Toronto), Giuseppe Pirrò (ICAR-CNR)

GraphGen: Exploring Interesting Graphs in Relational Data

Konstantinos Xirogiannopoulos (University of Maryland at College Park), Udayan Khurana (University of Maryland at College Park), Amol Deshpande (University of Maryland at College Park)

StarDB: A Large-Scale DBMS for Strings

Majed Sahli (King Abdullah University of Science and Technology), Essam Mansour (QCRI), Panos Kalnis (King Abdullah University of Science and Technology)

Wednesday Sep 2nd 18:00-21:00

Banquet

Location: Palace Lawn

Thursday Sep 3rd 07:30-08:30

Light Breakfast

Location: Grand Promenade

Thursday Sep 3rd 08:30-10:00

Award Ceremony: Award distribution for 10 year best paper, early career, and conference best papers

Location: Monarchy Ballroom

Chair: Surajit Chaudhuri

Award Ceremony

Thursday Sep 3rd 10:00-10:30

Refreshments: Coffee Break

Location: Grand Promenade

Thursday Sep 3rd 10:30-12:00

Research 21: Spatial Databases

Location: Kings 1

Chair: Walid Aref

Trajectory Simplification: On Minimizing the Direction-based Error

Cheng Long (Hong Kong University of Science and Technology), Raymond Chi-Wing Wong (Hong Kong University of Science and Technology), H V Jagadish (University of Michigan Ann Arbor)

Selectivity Estimation on Streaming SpatioTextual Data Using Local Correlations

Xiaoyang Wang (University of New South Wales), Ying Zhang (University of Technology (Sydney)), Wenjie Zhang (University of New South Wales), Xuemin Lin (University of New South Wales), Wei Wang (University of New South Wales)

Spatial Joins in Main Memory: Implementation Matters!

Dariusz Sidlauskas (Aarhus University), Christian Jensen (Aalborg University)

Large Scale Real-time Ridesharing with Service Guarantee on Road Networks

Yan Huang (University of North Texas), Favyen Bastani (MIT), Ruoming Jin (Kent State University), Xiaoyang Wang (Fudan University)

Compressed Spatial Hierarchical Bitmap (cSHB) Indexes for Efficiently Processing Spatial Range Query Workloads

Parth Nagarkar (Arizona State University), K. Selcuk Candan (Arizona State University), Aneesha Bhat (Arizona State University)

Research 22: Search

Location: Kings 2

Chair: Jennie Duggan

Finding Patterns in a Knowledge Base using Keywords to Compose Table Answers

Mohan Yang (UCLA), Bolin Ding (Microsoft Research), Surajit Chaudhuri (Microsoft Research), Kaushik Chakrabarti (Microsoft Research)

Searchlight: Enabling Integrated Search and Exploration over Large Multidimensional Data

Alexander Kalinin (Brown University), Ugur Cetintemel (Brown University), Stan Zdonik (Brown University)

Processing Moving kNN Queries Using Influential Neighbor Sets

Chuanwen Li (Northeastern University), Yu Gu (Northeastern University), Jianzhong Qi (University of Melbourne), Ge Yu (Northeastern University), Rui Zhang (University of Melbourne), Wang Yi (Northeastern University)

Reverse k Nearest Neighbors Query Processing: Experiments and Analysis

Shiyu Yang (University of New South Wales), Muhammad Cheema (Monash University), Xuemin Lin (University of New South Wales), Wei Wang (University of New South Wales)

Permutation Search Methods are Efficient, Yet Faster Search is Possible

Bilegsaikhan Naidan (NTNU), Leonid Boytsov (Carnegie Mellon University), Eric Nyberg (Carnegie Mellon University)

Industrial 6: Logging, Parallel Processing, and Graph Processing

Location: Kings 3

Chair: Bettina Kemme

Building a Replicated Logging System with Apache Kafka

Guozhang Wang (LinkedIn), Joel Koshy (LinkedIn), Sriram Subramanian (LinkedIn), Kartik Paramasivam (LinkedIn), Mammad Zadeh (LinkedIn), Neha Narkhede (Confluent Inc.), Jun Rao (Confluent Inc.), Jay Kreps (Confluent Inc.), Joe Stein (Big Data Open Source Security LLC)

Optimization of Common Table Expressions in MPP Database Systems

Amr El-Helw (Pivotal Inc.), Venkatesh Raghavan (Pivotal Inc.), Mohamed Soliman (Pivotal Inc.), George Caragea (Pivotal Inc.), Zhongxian Gu (Datometry Inc.), Michalis Petropoulos (Amazon Web Services)

One Trillion Edges: Graph Processing at Facebook-Scale

Avery Ching (Facebook), Dionysios Logothetis (Facebook), Sergey Edunov (Facebook), Maja Kabiljo (Facebook), Sambavi Muthukrishnan (Facebook)

Research 23: Logic and Semantics

Location: Queens 4

Chair: Tore Risch

DREAM: Distributed RDF Engine with Adaptive Query Planner and Minimal Communication

Mohammad Hammoud (Carnegie Mellon University), Dania Abed Rabbou (Carnegie Mellon University), Reza Nouri (University of New South Wales), Amin Beheshti (University of New South Wales), Sherif Sakr (University of New South Wales)

Efficient Identification of Implicit Facts in Incomplete OWL2-EL Knowledge Bases

John Liagouris (National Technical University of Athens), Manolis Terrovitis (IMIS `Athena')

Taming Subgraph Isomorphism for RDF Query Processing

Jinha Kim (Oracle Labs), Hyungyu Shin (POSTECH), Wook-Shin Han (POSTECH), Sungpack Hong (Oracle Labs), Hassan Chafi (Oracle Labs)

SEMA-JOIN : Joining Semantically-Related Tables Using Big Table Corpora

Yeye He (Microsoft Research), Kris Ganjam (Microsoft Research), Xu Chu (University of Waterloo)

QuickFOIL: Scalable Inductive Logic Programming

Qiang Zeng (University of Wisconsin-Madison), Jignesh Patel (University of Wisconsin-Madison), David Page (University of Wisconsin-Madison)

Research 24: Innovative Systems

Location: Queens 5

Chair: TBA

ScalaGiST: Scalable Generalized Search Trees for MapReduce Systems

Peng Lu (National University of Singapore), Gang Chen (Zhejiang University), Beng Chin Ooi (National University of Singapore), Hoang Tam Vo (National University of Singapore), Sai Wu (Zhejiang University)

DIADEM: Thousands of Websites to a Single Database

Tim Furche (Oxford University), Georg Gottlob (University of Oxford), Giovanni Grasso (Oxford University), Xiaonan Guo (Oxford University), Giorgio Orsi (University of Oxford), Christian Schallhart (Oxford University), Cheng Wang (Oxford University)

AsterixDB: A Scalable, Open Source BDMS

Sattam Alsubaiee (UC Irvine), Yasser Altowim (UC Irvine), Hotham Altwaijry (UC Irvine), Alex Behm (Cloudera), Vinayak Borkar (UC Irvine), Yingyi Bu (UC Irvine), Michael Carey (UC Irvine), Inci Cetindil (UC Irvine), Madhusudan Cheelangi (Google), Khurram Faraaz (IBM), Eugenia Gabrielova (UC Irvine), Raman Grover (UC Irvine), Zachary Heilbron (UC Irvine), Young-Seok Kim (UC Irvine), Chen Li (University of California (Irvine), Guangqiang Li (MarkLogic), Ji Mahn Ok (UC Irvine), Nicola Onose (Pivotal Inc.), Pouria Pirzadeh (UC Irvine), Vassilis Tsotras (UC Riverside), Rares Vernica (HP Labs), Jian Wen (Oracle Lab), Till Westmann (Oracle Labs)

Mega-KV: A Case for GPUs to Maximize the Throughput of In-Memory Key-Value Stores

Kai Zhang (University of Science and Technology of China), Kaibo Wang (The Ohio State University), Yuan Yuan (The Ohio State University), Lei Guo (The Ohio State University), Rubao Lee (The Ohio State University), Xiaodong Zhang (The Ohio State University)

UDA-GIST: An In-database Framework to Unify Data-Parallel and State-Parallel Analytics

Kun Li (University of Florida), Daisy Zhe Wang (University of Florida), Alin Dobra (University of Florida), Chris Dudley (University of Florida)

Tutorial 6: Truth Discovery and Crowdsourcing Aggregation: A Unified Perspective

Location: Queens 6

Truth Discovery and Crowdsourcing Aggregation: A Unified Perspective

Jing Gao, Qi Li, Bo Zhao, Wei Fan, Jiawei Han

Demo 2: Information Retrieval, Data Quality, and Provenance

Location: Kona 4

Chair: Tianli Mo

A Topic-based Reviewer Assignment System

Ngai Meng Kou (University of Macau), Leong Hou U (University of Macau), Nikos Mamoulis (University of Hong Kong), Yuhong Li (University of Macau), Ye Li (University of Macau), Zhiguo Gong (University of Macau)

Data Profiling with Metanome

Thorsten Papenbrock (Hasso-Plattner-Institute), Tanja Bergmann (Hasso-Plattner-Institute), Moritz Finke (Hasso-Plattner-Institute), Jakob Zwiener (Hasso-Plattner-Institute), Felix Naumann (Hasso-Plattner-Institute)

Provenance for SQL through Abstract Interpretation: Value-less, but Worthwhile

Tobias Müller (U Tübingen), Torsten Grust (U Tübingen)

SAASFEE: Scalable Scientific Workflow Execution Engine

Marc Bux (Humboldt-Universität zu Berlin), Jörgen Brandt (Humboldt-Universität zu Berlin), Carsten Lipka (Humboldt-Universität zu Berlin), Kamal Hakimzadeh (KTH Royal Institute of Technology), Jim Dowling (KTH Royal Institute of Technology), Ulf Leser (Humboldt Universität zu Berlin)

QOCO: A Query Oriented Data Cleaning System with Oracles

Moria Bergman (Tel Aviv University), Tova Milo (Tel Aviv University), Slava Novgorodov (Tel Aviv University), Wang-Chiew Tan (University of California Santa Cruz)

Collaborative Data Analytics with DataHub

Anant Bhardwaj (MIT), Amol Deshpande (University of Maryland), Aaron Elmore (University of Chicago), David Karger (MIT), Sam Madden (MIT), Aditya Parameswaran (University of Illinois at Urbana Champaign), Harihar Subramanyam (MIT), Eugene Wu (Columbia), Rebecca Zhang (MIT)

Mindtagger: A Demonstration of Data Labeling in Knowledge Base Construction

Jaeho Shin (Stanford University), Christopher Re (Stanford University), Mike Cafarella (University of Michigan)

Annotating Database Schemas to Help Enterprise Search

Eli Cortez (Microsoft), Philip Bernstein (Microsoft Research), Yeye He (Microsoft Research), Lev Novik (Microsoft)

KATARA: Reliable Data Cleaning with Knowledge Bases and Crowdsourcing

Xu Chu (University of Waterloo), John Morcos (University of Waterloo), Ihab Ilyas (University of Waterloo), Mourad Ouzzani (QCRI), Paolo Papotti (QCRI), Nan Tang (QCRI), Yin Ye (Google)

Gain Control over your Integration Evaluations Using iBench

Patricia Arocena (University of Toronto), Radu Ciucanu (University of Lille (INRIA), Boris Glavic (IIT), Renee Miller (University Toronto)

Janiform Intra-Document Analytics for Reproducible Research

Jens Dittrich (Saarland University), Patrick Bender (Saarland University)

EFQ: Why-Not Answer Polynomials in Action

Katerina Tzompanaki (Université Paris Sud), Nicole Bidoit (Université Paris Sud - INRIA), Melanie Herschel (University of Stuttgart)

Error Diagnosis and Data Profiling with Data X-Ray

Xiaolan Wang (University of Massachusetts Amherst), Mary Feng (University of Massachusetts Amherst and University of Iowa), Yue Wang (University of Massachusetts Amherst), Xin Luna Dong (Google Inc), Alexandra Meliou (University of Massachusetts Amherst)

Sharing and Reproducing Database Applications

Quan Pham (University of Chicago), Severin Thaler (University of Chicago), Tanu Malik (University of Chicago), Ian Foster (University of Chicago), Boris Glavic (IIT)

A Demonstration of TripleProv: Tracking and Querying Provenance over Web Data

Marcin Wylot (University of Fribourg), Philippe Cudré-Mauroux (University of Fribourg), Paul Groth (Elsevir Labs)

WADaR: Joint Wrapper and Data Repair

Stefano Ortona (University of Oxford), Giorgio Orsi (University of Oxford), Marcello Buoncristiano (Universita della Basilicata), Tim Furche (University of Oxford)

Wisteria: Nurturing Scalable Data Cleaning Infrastructure

Daniel Haas (UC Berkeley), Sanjay Krishnan (UC Berkeley), Jiannan Wang (UC Berkeley), Michael Franklin (UC Berkeley), Eugene Wu (Columbia University)

Thursday Sep 3rd 12:00-13:30

Lunch

Location: Lagoon Lanai

Thursday Sep 3rd 13:30-15:00

Research 25: Probabilistic Data Processing and Approximation

Location: Kings 1

Chair: Daisy Zhe Wang

Auto-Approximation of Graph Computing

Zechao Shang (Chinese University of Hong Kong), Jeffrey Xu Yu (Chinese University of Hong Kong)

Approximate lifted inference with probabilistic databases

Wolfgang Gatterbauer (Carnegie Mellon University), Dan Suciu (University of Washington)

Incremental Knowledge Base Construction Using DeepDive

Jaeho Shin (Stanford University), Sen Wu (Stanford University), Feiran Wang (Stanford University), Christopher De Sa (Stanford University), Ce Zhang (University of Wisconsin-Madison), Christopher Re (Stanford University)

Lenses: An On-Demand Approach to ETL

Ying Yang (SUNY Buffalo), Niccolò Meneghetti (SUNY Buffalo), Ronny Fehling (Oracle), Zhen Hua Liu (Oracle), Oliver Kennedy (SUNY Buffalo)

Knowledge-Based Trust: A Method to Estimate the Trustworthiness of Web Sources

Xin Luna Dong (Google Inc), Evgeniy Gabrilovich (Google Inc.), Kevin Murphy (Google Inc.), Van Dang (Google Inc.), Wilko Horn (Google Inc.), Camillo Lugaresi (Google Inc.), Shaohua Sun (Google Inc.), Wei Zhang (Google Inc.)*

Research 26: Query Processing 2

Location: Kings 2

Chair: Ken Salem

DAQ: A New Paradigm for Approximate Query Processing

Navneet Potti (University of Wisconsin-Madison), Jignesh Patel (University of Wisconsin-Madison)

On the Surprising Difficulty of Simple Things: the Case of Radix Partitioning

Felix Schuhknecht (Saarland University), Pankaj Khanchandani (Saarland University), Jens Dittrich (Saarland University)

Efficient Processing of Window Functions in Analytical SQL Queries

Viktor Leis (TU Munich), Kan Kundhikanjana (TU Munich), Alfons Kemper (TU Munich), Thomas Neumann (TU Munich)

Scaling Similarity Joins over Tree-Structured Data

Yu Tang (University of Hong Kong and EPFL), Yilun Cai (University of Hong Kong), Nikos Mamoulis (University of Hong Kong)

Processing of Probabilistic Skyline Queries Using MapReduce

Yoonjae Park (Seoul National University), Jun-Ki Min (Korea University of Technology and Education), Kyuseok Shim (Seoul National University)

Industrial 7: Privacy and Visualization

Location: Kings 3

Chair: Khan Arijit

Differential Privacy in Telco Big Data Platform

Xueyang Hu (Shanghai Jiao Tong University and Huawei Noah's Ark Lab), Mingxuan Yuan (Huawei Noah's Ark Lab), Jianguo Yao (Shanghai Jiao Tong University), Yu Deng (Shanghai Jiao Tong University), Lei Chen (Hong Kong University of Science and Technology), Haibing Guan (Shanghai Jiao Tong University), Jia Zeng (Soochow University and Huawei Noah's Ark Lab)

Efficient Evaluation of Object-Centric Exploration Queries for Visualization

You Wu (Duke University), Boulos Harb (Google Inc.), Jun Yang (Duke University), Cong Yu (Google Research)

Research 27: Data Warehousing, Search, and Ranking

Location: Queens 4

Chair: H.V. Jagadish

Interpretable and Informative Explanations of Outcomes

Kareem El Gebaly (University of Waterloo), Parag Agrawal (Twitter), Lukasz Golab (University of Waterloo), Filip Korn (Google), Divesh Srivastava (AT&T Labs-Research)

Stale View Cleaning: Getting Fresh Answers from Stale Materialized Views

Sanjay Krishnan (UC Berkeley), Jiannan Wang (UC Berkeley), Michael Franklin (UC Berkeley), Ken Goldberg (UC Berkeley), Tim Kraska (Brown University)

Scalable Topical Phrase Mining from Text Corpora

Ahmed El-Kishky (University of Illinois at Urbana Champaign), Yanglei Song (University of Illinois at Urbana Champaign), Chi Wang (Microsoft Research), Clare Voss (Army Research Laboratory), Jiawei Han (University of Illinois at Urbana Champaign)

Maximum Rank Query

Kyriakos Mouratidis (Singapore Management University), Jilian Zhang (Singapore Management University), HweeHwa Pang (Singapore Management University)

A Confidence-Aware Approach for Truth Discovery on Long-Tail Data

Qi Li (SUNY Buffalo), Yaliang Li (SUNY Buffalo), Jing Gao (SUNY Buffalo), Lu Su (SUNY Buffalo), Bo Zhao (Microsoft Research), Murat Demirbas (SUNY Buffalo), Wei Fan (Huawei Noah's Ark Lab), Jiawei Han (UIUC)

Research 28: Novel DB Architectures, Novel Hardware, and Resource Management

Location: Queens 5

Chair: Asterios Katsifodimos

An Architecture for Compiling UDF-centric Workflows

Andrew Crotty (Brown University), Alex Galakatos (Brown University), Kayhan Dursun (Brown University), Tim Kraska (Brown University), Carsten Binnig (Brown University), Ugur Cetintemel (Brown University), Stan Zdonik (Brown University)

Take me to your leader! Online Optimization of Distributed Storage Configurations

Artyom Sharov (Technion Israel Institute of Technology), Alexander Shraer (Google), Arif Merchant (Google), Murray Stokely (Google)

SIMD- and Cache-Friendly Algorithm for Sorting an Array of Structures

Hiroshi Inoue (IBM Research-Tokyo), Kenjiro Taura (University of Tokyo)

To Lock, Swap, or Elide: On the Interplay of Hardware Transactional Memory and Lock-Free Indexing

Darko Makreshanski (ETH Zurich), Justin Levandoski (Microsoft Research), Ryan Stutsman (Microsoft Research)

SQLite Optimization with Phase Change Memory for Mobile Applications

Gihwan Oh (Sungkyunkwan Univ), Sangchul Kim (Seoul National University), Sang-Won Lee (Sungkyunkwan University), Bongki Moon (Seoul National University)

Tutorial 7: Real Time Analytics: Algorithms and Systems (1/2)

Location: Queens 6

Real Time Analytics: Algorithms and Systems (1/2)

Arun Kejariwal, Sanjeev Kulkarni, Karthik Ramasamy

Panel 2: Designing for Interaction: Broadening our View of Working with Data

Location: Kona 1-2-3

Chair: Joseph M. Hellerstein

Designing for Interaction: Broadening our View of Working with Data

Azza Abouzied (NYU-AD), Adam Marcus (Unlimited Labs), Arnab Nandi (Ohio State University), Eugene Wu (Columbia University), Joseph M. Hellerstein (UC Berkeley)



Bio: Azza Abouzied's research work focuses on designing intuitive data querying tools. Today's technologies are helping people collect and produce data at phenomenal rates. Despite the abundance of data, it remains largely inaccessible due to the skill required to explore, query and analyze it in a non-trivial fashion. While many users know exactly what they are looking for, they have trouble expressing sophisticated queries in interfaces that require knowledge of a programming language or a query language. Azza designs novel interfaces, such as example-driven query tools, that simplify data querying and analysis. Her research work combines techniques from various research fields such as UI-design, machine learning and databases. Azza Abouzied received her doctoral degree from Yale in 2013. She spent a year as a visiting scholar at UC Berkeley. She is also one of the co-founders of Hadapt – a Big Data analytics platform.



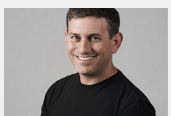
Bio: Adam just cofounded Unlimited Labs, a company dedicated to the future of work. Prior to that, Adam led the data team at Locu, a startup that was acquired by GoDaddy. He completed his Ph.D. in computer science at MIT in 2012, where his dissertation was on database systems and human computation. Adam is a recipient of the NSF and NDSEG fellowships, and has previously worked at ITA, Google, IBM, and FactSet. In his free time, Adam builds course content to get people excited about data and programming.



Bio: Arnab's research is in the area of database systems, focusing on challenges in big data analytics and interactive query interfaces. The goal of his group is to empower humans to effectively interact with data. This involves solving problems that span the areas of databases, visualization, human-computer interaction, and information retrieval.



Bio: Eugene is broadly interested in technologies that help users play with their data. His goal is for users at all technical levels to effectively and quickly make sense of their information. He is interested in solutions that ultimately improve the interface between users and data, and techniques borrows from fields such as data management, systems, crowd sourcing, visualization, and HCI. Eugene is starting at Columbia University in Fall of 2015.



Bio: Joe is a Professor of Computer Science at the University of California, Berkeley, whose work focuses on data-centric systems and the way they drive computing. He is an ACM Fellow, an Alfred P. Sloan Research Fellow and the recipient of three SIGMOD "Test of Time" awards for his research. In 2010, Fortune Magazine included him in their list of 50 smartest people in technology, and MIT's Technology Review magazine included his work on their TR10 list of the 10 technologies "most likely to change our world". Joe is also the co-founder and Chief Strategy Officer of Trifacta, a software vendor providing intelligent interactive solutions to the messy problems of wrangling data. He serves on the technical advisory boards of a number of computing and Internet companies including EMC, SurveyMonkey, Captricity, and Dato, and previously served as the Director of Intel Research, Berkeley.

Thursday Sep 3rd 15:00-15:30

Refreshments: Coffee Break

Location: Grand Promenade

Thursday Sep 3rd 15:30-17:00

Research 29: Privacy and Security

Location: Kings 1

Chair: Johann Christoph Freytag

Practical Authenticated Pattern Matching with Optimal Proof Size

Dimitrios Papadopoulos (Boston University), Charalampos Papamanthou (University of Maryland), Roberto Tamassia (Brown University), Nikos Triandopoulos (RSA Laboratories and Boston University)

Fast Range Query Processing with Strong Privacy Protection for Cloud Computing

Rui Li (Hunan university), Alex Liu (Michigan State University), Ann Wang (Michigan State University), Bezawada Bruhadeshwar (Nanjing University)

DPT: Differentially Private Trajectory Synthesis Using Hierarchical Reference Systems

Xi He (Duke University), Graham Cormode (University of Warwick), Ashwin Machanavajjhala (Duke University), Cecilia Procopiuc (AT&T Labs-Research), Divesh Srivastava (AT&T Labs-Research)

Privacy Implications of Database Ranking

Md Farhadur Rahman (University of Texas at Arlington), Weimo Liu (George Washington University), Saravanan Thirumuruganathan (University of Texas at Arlington), Nan Zhang (George Washington University), Gautam Das (University of Texas at Arlington)

Research 30: Logic Programming, Web Data Management and Query Processing

Location: Kings 2

Chair: Yannis Papakonstantinou

Selective Provenance for Datalog Programs Using Top-k Queries

Daniel Deutch (Tel Aviv University), Amir Gilad (Tel Aviv University), Yuval Moskovitch (Tel Aviv University)

Asynchronous and Fault-Tolerant Recursive Datalog Evaluation in Shared-Nothing Engines

Jingjing Wang (University of Washington), Magdalena Balazinska (University of Washington), Daniel Halperin (University of Washington)

Aggregate Estimations over Location Based Services

Weimo Liu (George Washington University), Md Farhadur Rahman (University of Texas at Arlington), Saravanan Thirumuruganathan (University of Texas at Arlington), Nan Zhang (George Washington University), Gautam Das (University of Texas at Arlington)

PARADIS: An Efficient Parallel Algorithm for In-place Radix Sort

Minsik Cho (IBM T.J. Watson Research Center), Daniel Brand (IBM T.J. Watson Research Center), Rajesh Bordawekar (IBM T.J. Watson Research Center), Ulrich Finkler (IBM T.J. Watson Research Center), Vincent Kulkandaisamy (IBM T.J. Watson Research Center), Ruchir Puri (IBM T.J. Watson Research Center)

Performance and Scalability of Indexed Subgraph Query Processing Methods

Foteini Katsarou (University of Glasgow), Nikos Ntarmos (University of Glasgow), Peter Triantafillou (University of Glasgow)

Research 31: Structure and Dependency Discovery, and Spatial Databases

Location: Kings 3

Chair: Ziawasch Abedjan

Spatial Partitioning Techniques in SpatialHadoop

Ahmed Eldawy (University of Minnesota), Louai Alarabi (University of Minnesota), Mohamed Mokbel (University of Minnesota)

Divide & Conquer-based Inclusion Dependency Discovery

Thorsten Papenbrock (Hasso-Plattner-Institute), Sebastian Kruse (Hasso-Plattner-Institute), Jorge-Arnulfo Quijano-Ruiz (Qatar Research Institute), Felix Naumann (Hasso Plattner Institut Potsdam)

Functional Dependency Discovery: An Experimental Evaluation of Seven Algorithms

Thorsten Papenbrock (Hasso-Plattner-Institute), Jens Ehrlich (Hasso-Plattner-Institute), jannik Marten (Hasso-Plattner-Institute), Tommy Neubert (Hasso-Plattner-Institute), Jan-Peer Rudolph (Hasso-Plattner-Institute), Martin Schönberg (Hasso-Plattner-Institute), Jakob Zwiener (Hasso-Plattner-Institute), Felix Naumann (Hasso-Plattner-Institute)

Extracting Logical Hierarchical Structure of HTML Documents Based on Headings

Tomohiro Manabe (Kyoto University), Keishi Tajima (Kyoto University)

Bonding Vertex Sets Over Distributed Graph: A Betweenness Aware Approach

Xiaofei Zhang (HKSUT), Hong Cheng (The Chinese University of Hong Kong), Lei Chen (Hong Kong University of Science and Technology)

Research 32: Time Series and Streams

Location: Queens 4

Chair: Themis Palpanis

CANDS: Continuous Optimal Navigation via Distributed Stream Processing

Dingyu Yang (Shanghai Jiao Tong University), Dongxiang Zhang (National University of Singapore), Kian-Lee Tan (National University of Singapore), Jian Cao (Shanghai Jiao Tong University), Frédéric Le Mouël (University of Lyon)

General Incremental Sliding-Window Aggregation

Kanat Tangwongsan (Mahidol University International College), Martin Hirzel (IBM T. J. Watson Research Center), Scott Schneider (IBM T. J. Watson Research Center), Kun-Lung Wu (IBM T.J. Watson Research Center)

YADING: Fast Clustering of Large-Scale Time Series Data

Rui Ding (Microsoft Research), Qiang Wang (Microsoft Research), Yingnong Dang (Microsoft Research), Qiang Fu (Microsoft Research), Haidong Zhang (Microsoft Research), Dongmei Zhang (Microsoft Research)

Monitoring Distributed Streams using Convex Decompositions

Arnon Lazerson (Technion), Daniel Keren (Haifa University), Izchak Sharfman (Technion), Minos Garofalakis (Technical University of Crete), Vasilis Samoladas (Technical University of Crete), Assaf Schuster (Technion)

Research 33: Transaction Processing

Location: Queens 5

Chair: Heiko Schuldt

Staring into the Abyss: An Evaluation of Concurrency Control with One Thousand Cores

Xiangyao Yu (MIT), George Bezerra (MIT), Andy Pavlo (Carnegie Mellon University), Srinivas Devadas (MIT), Michael Stonebraker (MIT)

E-Store: Fine-Grained Elastic Partitioning for Distributed Transaction Processing Systems

Rebecca Taft (MIT), Essam Mansour (QCRI), Marco Serafini (QCRI), Jennie Duggan (MIT), Aaron Elmore (MIT), Ashraf Aboulnaga (QCRI), Andy Pavlo (Carnegie Mellon University), Michael Stonebraker (MIT)

ADDICT: Advanced Instruction Chasing for Transactions

Pinar Tozun (EPFL), Islam Atta (University of Toronto), Anastasia Ailamaki (EPFL), Andrea Moshovos (University of Toronto)

Rethinking serializable multiversion concurrency control

Jose Faleiro (Yale University), Daniel Abadi (Yale University)

Tutorial 8: Real Time Analytics: Algorithms and Systems (2/2)

Location: Queens 6

Real Time Analytics: Algorithms and Systems (2/2)

Arun Kejariwal, Sanjeev Kulkarni, Karthik Ramasamy

Demo 3: Systems, User Interfaces, and Visualization

Location: Kona 4

Chair: Tianli Mo

FP-Hadoop: Efficient Execution of Parallel Jobs Over Skewed Data

Miguel Liroz-Gistau (INRIA), Reza Akbarinia (INRIA), Patrick Valduriez (INRIA)

SDB: A Secure Query Processing System with Data Interoperability

Zhian He (Hong Kong Polytechnic University), WaiKit Wong (Hang Seng Management College), Ben Kao (University of Hong Kong), David W. Cheung (University of Hong Kong), Rongbin Li (University of Hong Kong), Siu Ming Yiu (University of Hong Kong), Eric Lo (Polytechnic University of Hong Kong)

A Demonstration of HadoopViz: An Extensible MapReduce-based System for Visualizing Big Spatial Data

Ahmed Eldawy (University of Minnesota), Mohamed Mokbel (University of Minnesota), Christopher Jonathan (University of Minnesota)

A Demonstration of the BigDAWG Multi-Database System

Aaron Elmore (MIT), Jennie Duggan (Northwestern), Michael Stonebraker (MIT), Manasi Vartak (MIT), Sam Madden (MIT), Vijay Gadepally (MIT), Jeremy Kepner (MIT), Timothy Mattson (Intel), Jeff Parhurst (Intel), Stavros Papadopoulos (MIT), Nesime Tatbul (Intel Labs and MIT), Magdalena Balazinska (University of Washington), Bill Howe (University of Washington), Jeffrey Heer (University of Washington), David Maier (Portland State University), Tim Kraska (Brown), Ugur Cetintemel (Brown University), Stan Zdonik (Brown University)

RINSE: Interactive Data Series Exploration

Kostas Zoumpatianos (University of Trento), Stratos Idreos (Harvard), Themis Palpanas (Paris Descartes University)

Smart Drill-Down: A New Data Exploration Operator

Manas Joglekar (Stanford University), Hector Garcia-Molina (Stanford University), Aditya Parameswaran (University of Illinois at Urbana Champaign)

VIQ: auto-suggestion enabled visual interface for interactive graph query formulation

Nandish Jayaram (University of Texas at Arlington), Sidharth Goyal (University of Texas at Arlington), Chengkai Li (University of Texas at Arlington)

VINERy: A Visual IDE for Information Extraction

Yunyao Li (IBM Research-Almaden), Elmer Kim (Treasurer Data (Inc.)), Marc Touchette (IBM Silicon Valley Lab), Ramiya Venkatachalam (IBM Silicon Valley Lab), Hao Wang (IBM Silicon Valley Lab)

GIS navigation boosted by column stores

Foteini Alvanaki (CWI), Romulo Goncalves (Netherlands eScience Center), Milena Ivanova (NuoDB), Martin Kersten (CWI), Kostis Kyzirakos (CWI)

AIDE: An Automatic User Navigation System for Interactive Data Exploration

Yanlei Diao (University of Massachusetts Amherst), Kyriaki Dimitriadou (Brandeis university), Zhan Li (Brandeis University), Wenzhao Liu (University of Massachusetts Amherst), Olga Papaemmanouil (Brandeis University), Kemi Peng (Brandeis University), Liping Peng (University of Massachusetts Amherst)

A Demonstration of AQWA: Adaptive Query-Workload-Aware Partitioning of Big Spatial Data

Ahmed Aly (Purdue University), Ahmed Abdelhamid (Purdue University), Ahmed Mahmood (Purdue University), Walid Aref (Purdue University), Mohamed Hassan (Purdue University), Hazem Elmeleegy (Turn Inc), Mourad Ouzzani (Qatar Computing Research Institute)

Data-Spread: Unifying Databases and Spreadsheets

Mangesh Bendre (University of Illinois at Urbana-Champaign), Bofan Sun (University of Illinois at Urbana-Champaign), Ding Zhang (University of Illinois at Urbana-Champaign), Xinyan Zhou (University of Illinois at Urbana-Champaign), Kevin Chang (University of Illinois at Urbana-Champaign), Aditya Parameswaran (University of Illinois at Urbana-Champaign)

CODD: A Dataless Approach to Big Data Testing

Ashoke S (Indian Institute of Science), Jayant Haritsa (IISc)

Vizdom: Interactive Analytics through Pen and Touch

Andrew Crotty (Brown University), Alex Galakatos (Brown University), Emanuel Zraggen (Brown University), Carsten Binnig (Brown University), Tim Kraska (Brown University)

DBSeer: Pain-free Database Administration through Workload Intelligence

Dong Young Yoon (University of Michigan Ann Arbor), Barzan Mozafari (University of Michigan Ann Arbor), Douglas Brown (Teradata Inc.)

Thursday Sep 3rd 17:00-19:00

Reception and Poster Session 2

Location: Kohala Ballroom

Chair: Tianli Mo

TOP: A Framework for Enabling Algorithmic Optimizations for Distance-Related Problems

Yufei Ding - Xipeng Shen - Madanlal Musuvathi - Todd Mytkowicz

SCAN++: Efficient Algorithm for Finding Clusters, Hubs and Outliers on Large-scale Graphs

Hiroaki Shiokawa - Yasuhiro Fujiwara - Makoto Onizuka

GraphTwist: Fast Iterative Graph Computation with Two-tier Optimizations

Yang Zhou - Ling Liu - Kisung Lee - Qi Zhang

A Scalable Distributed Graph Partitioner

Daniel Margo - Margo Seltzer

Keys for Graphs

Wenfei Fan - Zhe Fan - Chao Tian - Xin Luna Dong

Scaling Up Concurrent Main-Memory Column-Store Scans: Towards Adaptive NUMA-aware Data and Task Placement

Iraklis Psaroudakis - Tobias Scheuer - Norman May - Abdelkader Sellami - Anastassia Ailamaki

In-Memory Performance for Big Data

Goetz Graefe - Haris Volos - Hideaki Kimura - Harumi Kuno - Joseph Tucek - Mark Lillibridge - Alistair Veitch

Profiling R on a Contemporary Processor

Shriram Sridharan - Jignesh Patel

Deployment of Query Plans on Multicores

Jana Giceva - Gustavo Alonso - Timothy Roscoe - Tim Harris

Faster Set Intersection with SIMD instructions by Reducing Branch Mispredictions

Hiroshi Inoue - Moriyoshi Ohara - Kenjiro Taura

Resource Bricolage for Parallel Database Systems

Jiexing Li - Jeffrey Naughton - Rimma Nehme

Multi-Objective Parametric Query Optimization

Immanuel Trummer - Christoph Koch

Querying with Access Patterns and Integrity Constraints

Michael Benedikt - Julien Leblay - Efi Tsamoura

Uncertainty Aware Query Execution Time Prediction

Wentao Wu - Xi Wu - Hakan Hacigumus - Jeffrey Naughton

Join Size Estimation Subject to Filter Conditions

David Vengerov - Andre Menck - Sunil Chakkappen - Mohamed Zait

Scaling Up Crowd-Sourcing to Very Large Datasets: A Case for Active Learning

Barzan Mozafari - Purna Sarkar - Michael Franklin - Michael Jordan - Sam Madden

Hear the Whole Story: Towards the Diversity of Opinion in Crowdsourcing Markets

Ting Wu - Lei Chen - Pan Hui - CHEN ZHANG - Weikai Li

Where To: Crowd-Aided Path Selection

Chen Zhang - Yongxin Tong - Lei Chen

Reliable Diversity-Based Spatial Crowdsourcing by Moving Workers

Peng Cheng - Xiang Lian - Zhao Chen - Rui Fu - Lei Chen - Jinsong Han - Jizhong Zhao

Learning User Preferences By Adaptive Pairwise Comparison

Li Qian - Jinyang Gao - H V Jagadish

Pregelx: Big(ger) Graph Analytics on A Dataflow Engine

Yingyi Bu - Vinayak Borkar - Jianfeng Jia - Michael Carey - Tyson Condie

Large-Scale Distributed Graph Computing Systems: An Experimental Evaluation

Yi Lu - James Cheng - Da Yan - Huanhuan Wu

Fast Failure Recovery in Distributed Graph Processing Systems

Yanyan Shen - Gang Chen - H V Jagadish - Wei Lu - Beng Chin Ooi - Bogdan Tudor

Giraph Unchained: Barrierless Asynchronous Parallel Execution in Pregel-like Graph Processing Systems

Minyang Han - Khuzaima Daudjee

GraphMat: High performance graph analytics made productive

Narayanan Sundaram - Nadathur Satish - Mostofa Ali Patwary - Subramanya Dulloor - Michael Anderson - Satya Gautam Vadlamudi - Dipankar Das - Pradeep Dubey

In-Cache Query Co-Processing on Coupled CPU-GPU Architectures

Jiong He - Shuhao Zhang - Bingsheng He

NVRAM-aware Logging in Transaction Systems

Jian Huang - Karsten Schwan - Moinuddin Qureshi

Improving Main Memory Hash Joins on Intel Xeon Phi Processors: An Experimental Approach

Saurabh Jha - Bingsheng He - Mian Lu - Xuntao Cheng - Phung Huynh Huynh

REWIND: Recovery Write-Ahead System for In-Memory Non-Volatile Data-Structures

Andreas Chatzistergiou - Marcelo Cintra - Stratis Viglas

Persistent B+-Trees in Non-Volatile Main Memory

Shimin Chen - Qin Jin

Robust Local Community Detection: On Free Rider Effect and Its Elimination
Yubao Wu - Ruoming Jin - Jing Li - Xiang Zhang

Viral Marketing Meets Social Advertising: Ad Allocation with Minimum Regret
Cigdem Aslay - Wei Lu - Francesco Bonchi - Amit Goyal - Laks Lakshmanan

Community Detection in Social Networks: An In-depth Benchmarking Study with a Procedure-Oriented Framework
Meng Wang - Chaokun Wang - Jeffrey Xu Yu - Jun Zhang

Leveraging History for Faster Sampling of Online Social Networks
Zhuojie Zhou - Nan Zhang - Gautam Das

Real-time Targeted Influence Maximization for Online Advertisements
Yuchen Li - Dongxiang Zhang - Kian-Lee Tan

Top-k Nearest Neighbor Search In Uncertain Data Series
Michele Dallachiesa - Themis Palpanas - Ihab Ilyas

Scaling Manifold Ranking Based Image Retrieval
Yasuhiro Fujiwara - Go Irie - Shari Kuroyama - Makoto Onizuka

Optimal Enumeration: Efficient Top-k Tree Matching
Lijun Chang - Xuemin Lin - Wenjie Zhang - Jeffrey Xu Yu - Ying Zhang - Lu Qin

Generating Top-k Packages via Preference Elicitation
Min Xie - Laks V. S. Lakshmanan - Peter Wood

Rank aggregation with ties: Experiments and Analysis
Bryan Brancotte - Bo Yang - Guillaume Blin - Sarah Cohen-Boulakia - Alain Denise - Sylvie Hamel

Trajectory Simplification: On Minimizing the Direction-based Error
Cheng Long - Raymond Chi-Wing Wong - H V Jagadish

Selectivity Estimation on Streaming SpatioTextual Data Using Local Correlations
Xiaoyang Wang - Ying Zhang - Wenjie Zhang - Xuemin Lin - Wei Wang

Spatial Joins in Main Memory: Implementation Matters!
Darius Sidlauskas - Christian Jensen

Large Scale Real-time Ridesharing with Service Guarantee on Road Networks
Yan Huang - Favyen Bastani - Ruoming Jin - Xiaoyang Wang

Compressed Spatial Hierarchical Bitmap (cSHB) Indexes for Efficiently Processing Spatial Range Query Workloads
Parth Nagarkar - K. Selcuk Candan - Aneesha Bhat

Finding Patterns in a Knowledge Base using Keywords to Compose Table Answers
Mohan Yang - Bolin Ding - Surajit Chaudhuri - Kaushik Chakrabarti

Searchlight: Enabling Integrated Search and Exploration over Large Multidimensional Data
Alexander Kalinin - Ugur Cetintemel - Stan Zdonik

Processing Moving kNN Queries Using Influential Neighbor Sets
Chuanwen Li - Yu Gu - Jianzhong Qi - Ge Yu - Rui Zhang - Wang Yi

Reverse k Nearest Neighbors Query Processing: Experiments and Analysis
Shiyu Yang - Muhammad Cheema - Xuemin Lin - Wei Wang

Permutation Search Methods are Efficient, Yet Faster Search is Possible
Bilegsaikhan Naidan - Leonid Boytsov - Eric Nyberg

DREAM: Distributed RDF Engine with Adaptive Query Planner and Minimal Communication
Mohammad Hammoud - Dania Abed Rabbou - Reza Nouri - Seyed-Mehdi-Reza Beheshti - Sherif Sakr

Efficient Identification of Implicit Facts in Incomplete OWL2-EL Knowledge Bases
John Liagouris - Manolis Terrovitis

Taming Subgraph Isomorphism for RDF Query Processing
Jinha Kim - Hyungyu Shin - Wook-Shin Han - Sungpack Hong - Hassan Chafi

SEMA-JOIN : Joining Semantically-Related Tables Using Big Table Corpora
Yeye He - Kris Ganjam - Xu Chu

QuickFOIL: Scalable Inductive Logic Programming
Qiang Zeng - Jignesh Patel - David Page

ScalaGiST: Scalable Generalized Search Trees for MapReduce Systems
Peng Lu - Gang Chen - Beng Chin Ooi - Hoang Tam Vo - Sai Wu

DIADEM: Thousands of Websites to a Single Database
Tim Furche - Georg Gottlob - Giovanni Grasso - Xiaonan Guo - Giorgio Orsi - Christian Schallhart - Cheng Wang

AsterixDB: A Scalable, Open Source BDMS

Sattam Alsubaiee - Yasser Altowim - Hotham Altwajry - Alex Behm - Vinayak Borkar - Yingyi Bu - Michael Carey - Inci Cetindil - Madhusudan Cheelangi - Khurram Faraaz - Eugenia Gabrielova - Raman Grover - Zachary Heilbron - Young-Seok Kim - Chen Li - Guangqiang Li - Ji Mahn Ok - Nicola Onose - Pouria Pirzadeh - Vassilis Tsotras - Rares Vernica - Jian Wen - Till Westmann

Mega-KV: A Case for GPUs to Maximize the Throughput of In-Memory Key-Value Stores

Kai Zhang - Kaibo Wang - Yuan Yuan - Lei Guo - Rubao Lee - Xiaodong Zhang

UDA-GIST: An In-database Framework to Unify Data-Parallel and State-Parallel Analytics

Kun Li - Daisy Zhe Wang - Alin Dobra - Chris Dudley

Auto-Approximation of Graph Computing

Zechao Shang - Jeffrey Xu Yu

Approximate lifted inference with probabilistic databases

Wolfgang Gatterbauer - Dan Suciu

Incremental Knowledge Base Construction Using DeepDive

Jaeho Shin - Sen Wu - Feiran Wang - Christopher De Sa - Ce Zhang - Christopher Re

Lenses: An On-Demand Approach to ETL

Ying Yang - Niccolo Meneghetti - Ronny Fehling - Zhen Hua Liu - Oliver Kennedy

Knowledge-Based Trust: A Method to Estimate the Trustworthiness of Web Sources

Xin Luna Dong - Evgeniy Gabrilovich - Kevin Murphy - Van Dang - Wilko Horn - Camillo Lugaresi - Shaohua Sun - Wei Zhang

DAQ: A New Paradigm for Approximate Query Processing

Navneet Potti - Jignesh Patel

On the Surprising Difficulty of Simple Things: the Case of Radix Partitioning

Felix Schuhknecht - Pankaj Khanchandani - Jens Dittrich

Efficient Processing of Window Functions in Analytical SQL Queries

Viktor Leis - Kan Kundhikanjana - Alfons Kemper - Thomas Neumann

Scaling Similarity Joins over Tree-Structured Data

Yu Tang - Yilun Cai - Nikos Mamoulis

Processing of Probabilistic Skyline Queries Using MapReduce

Yoonjae Park - Jun-Ki Min - Kyuseok Shim

Interpretable and Informative Explanations of Outcomes

Kareem El Gebaly - Parag Agrawal - Lukasz Golab - Filip Korn - Divesh Srivastava

Stale View Cleaning: Getting Fresh Answers from Stale Materialized Views

Sanjay Krishnan - Jiannan Wang - Michael Franklin - Ken Goldberg - Tim Kraska

Scalable Topical Phrase Mining from Text Corpora

Ahmed El-Kishky - Yanglei Song - Chi Wang - Clare Voss - Jiawei Han

Maximum Rank Query

Kyriakos Mouratidis - Jilian Zhang - HweeHwa Pang

A Confidence-Aware Approach for Truth Discovery on Long-Tail Data

Qi Li - Yaliang Li - Jing Gao - Lu Su - Bo Zhao - Murat Demirbas - Wei Fan - Jiawei Han

An Architecture for Compiling UDF-centric Workflows

Andrew Crotty - Alex Galakatos - Kayhan Dursun - Tim Kraska - Carsten Binnig - Ugur Cetintemel - Stan Zdonik

Take me to your leader! Online Optimization of Distributed Storage Configurations

Artyom Sharov - Alexander Shraer - Arif Merchant - Murray Stokely

SIMD- and Cache-Friendly Algorithm for Sorting an Array of Structures

Hiroshi Inoue - Kenjiro Taura

To Lock, Swap, or Elide: On the Interplay of Hardware Transactional Memory and Lock-Free Indexing

Darko Makreshanski - Justin Levandoski - Ryan Stutsman

SQLite Optimization with Phase Change Memory for Mobile Applications

Gihwan Oh - Sangchul Kim - Sang-Won Lee - Bongki Moon

Practical Authenticated Pattern Matching with Optimal Proof Size

Dimitrios Papadopoulos - Charalampos Papamanthou - Roberto Tamassia - Nikos Triandopoulos

Fast Range Query Processing with Strong Privacy Protection for Cloud Computing

Rui Li - Alex Liu - Ann Wang - Bezawada Bruhadeshwar

DPT: Differentially Private Trajectory Synthesis Using Hierarchical Reference Systems

Xi He - Graham Cormode - Ashwin Machanavajjhala - Cecilia Procopiuc - Divesh Srivastava

Privacy Implications of Database Ranking

Md Farhadur Rahman - Weimo Liu - Saravanan Thirumuruganathan - Nan Zhang - Gautam Das

Selective Provenance for Datalog Programs Using Top-k Queries

Daniel Deutch - Amir Gilad - Yuval Moskovitch

Asynchronous and Fault-Tolerant Recursive Datalog Evaluation in Shared-Nothing Engines

Jingjing Wang - Magdalena Balazinska - Daniel Halperin

Aggregate Estimations over Location Based Services

Weimo Liu - Md Farhadur Rahman - Saravanan Thirumuruganathan - Nan Zhang - Gautam Das

PARADIS: An Efficient Parallel Algorithm for In-place Radix Sort

Minsik Cho - Daniel Brand - Rajesh Bordawekar - Ulrich Finkler - Vincent Kalandaisamy - Ruchir Puri

Performance and Scalability of Indexed Subgraph Query Processing Methods

Foteini Katsarou - Nikos Ntarmos - Peter Triantafillou

Spatial Partitioning Techniques in SpatialHadoop

Ahmed Eldawy - Louai Alarabi - Mohamed Mokbel

Divide & Conquer-based Inclusion Dependency Discovery

Thorsten Papenbrock - Sebastian Kruse - Jorge-Arnulfo Quijane-Ruiz - Felix Naumann

Functional Dependency Discovery: An Experimental Evaluation of Seven Algorithms

Thorsten Papenbrock - Jens Ehrlich - jannik Marten - Tommy Neubert - Jan-Peer Rudolph - Martin Sch`nberg - Jakob Zwiener - Felix Naumann

Extraction of Logical Structure of Documents Based on Hierarchical Headings

Tomohiro Manabe - Keishi Tajima

Bonding Vertex Sets Over Distributed Graph: A Betweenness Aware Approach

Xiaofei Zhang - Hong Cheng - Lei Chen

CANDS: Continuous Optimal Navigation via Distributed Stream Processing

Dingyu Yang - Dongxiang Zhang - Kian-Lee Tan - Jian Cao - Frederic Le Mouel

General Incremental Sliding-Window Aggregation

Kanat Tangwongsan - Martin Hirzel - Scott Schneider - Kun-Lung Wu

YADING: Fast Clustering of Large-Scale Time Series Data

Rui Ding - Qiang Wang - Yingnong Dang - Qiang Fu - Haidong Zhang - Dongmei Zhang

Monitoring Distributed Streams using Convex Decompositions

Arnon Lazerson - Daniel Keren - Izchak Sharfman - Minos Garofalakis - Vasilis Samoladas - Assaf Schuste

Staring into the Abyss: An Evaluation of Concurrency Control with One Thousand Cores

Xiangyao Yu - George Bezerra - Andy Pavlo - Srinivas Devadas - Michael Stonebraker

E-Store: Fine-Grained Elastic Partitioning for Distributed Transaction Processing Systems

Rebecca Taft - Essam Mansour - Marco Serafini - Jennie Duggan - Aaron Elmore - Ashraf Aboulnaga - Andy Pavlo - Michael Stonebraker

ADDICT: Advanced Instruction Chasing for Transactions

Pinar Tozun - Islam Atta - Anastasia Ailamaki - Andrea Moshovos

Rethinking serializable multiversion concurrency control

Jose Faleiro - Daniel Abadi

Friday Sep 4th 08:00-09:00

Light Breakfast

Location: Grand Promenade

Friday Sep 4th 09:00-10:30

DMAH Session 1

Location: Kings 1

First International Workshop on Data Management and Analytics for Medicine and Healthcare (DMAH 2015)

Fusheng Wang (Stony Brook University), Gang Luo (University of Utah), Chunhua Weng (Columbia University)

BPOE Session 1
Location: Kings 2

Sixth workshop on Big Data Benchmarks, Performance Optimization, and Emerging Hardware (BPOE-6)
Jianfeng Zhan (Chinese Academy of Sciences), Roberto V. Zicari (Goethe University), Rui Han (Chinese Academy of Sciences)

BOSS Session 1
Location: Queens 4-5-6

First workshop on Big Data Open Source System (BOSS 2015)
Tilman Rabl, TU Berlin

Friday Sep 4th 10:30-11:00

Refreshments: Coffee Break
Location: Grand Promenade

Friday Sep 4th 11:00-12:30

DMAH Session 2
Location: Kings 1

First International Workshop on Data Management and Analytics for Medicine and Healthcare (DMAH 2015)
Fusheng Wang (Stony Brook University), Gang Luo (University of Utah), Chunhua Weng (Columbia University)

BPOE Session 2
Location: Kings 2

Sixth workshop on Big Data Benchmarks, Performance Optimization, and Emerging Hardware (BPOE-6)
Jianfeng Zhan (Chinese Academy of Sciences), Roberto V. Zicari (Goethe University), Rui Han (Chinese Academy of Sciences)

BOSS Session 2
Location: Queens 4-5-6

First workshop on Big Data Open Source System (BOSS 2015)
Tilman Rabl, TU Berlin

Friday Sep 4th 12:30-14:00

Lunch
Location: Lagoon Lanai

Friday Sep 4th 14:00-15:00

DMAH Session 3
Location: Kings 1

First International Workshop on Data Management and Analytics for Medicine and Healthcare (DMAH 2015)
Fusheng Wang (Stony Brook University), Gang Luo (University of Utah), Chunhua Weng (Columbia University)

BPOE Session 3
Location: Kings 2

Sixth workshop on Big Data Benchmarks, Performance Optimization, and Emerging Hardware (BPOE-6)
Jianfeng Zhan (Chinese Academy of Sciences), Roberto V. Zicari (Goethe University), Rui Han (Chinese Academy of Sciences)

BOSS Session 3
Location: Queens 4-5-6

First workshop on Big Data Open Source System (BOSS 2015)

Tilmann Rabl, TU Berlin

Friday Sep 4th 15:00-15:30

Refreshments: Coffee Break

Location: Grand Promenade

Friday Sep 4th 15:30-18:00

DMAH Session 4

Location: Kings 1

First International Workshop on Data Management and Analytics for Medicine and Healthcare (DMAH 2015)

Fusheng Wang (Stony Brook University), Gang Luo (University of Utah), Chunhua Weng (Columbia University)

BPOE Session 4

Location: Kings 2

Sixth workshop on Big Data Benchmarks, Performance Optimization, and Emerging Hardware (BPOE-6)

Jianfeng Zhan (Chinese Academy of Sciences), Roberto V. Zicari (Goethe University), Rui Han (Chinese Academy of Sciences)

BOSS Session 4

Location: Queens 4-5-6

First workshop on Big Data Open Source System (BOSS 2015)

Tilmann Rabl, TU Berlin



IT'S NOT WHAT YOU LOOK AT.
IT'S WHAT YOU SEE!
RETHINK BIG DATA

- Reduced TCO
 - Real time analytics
 - Industry specific solutions
-

Infosys® | Information Platform

For more information, contact askus@infosys.com

The In-Memory Database Research Leader

SAP database and technology portfolio opens a completely new and exciting research field, and provides the best opportunity to attract the world's attention.

Join us to help the world run simpler and improve people's lives.

Find out about SAP In-Memory Database Solutions at saphana.com.

Get started at sap.com/careers

The SAP logo is located in the bottom left corner of the image. It consists of the letters 'SAP' in a bold, white, sans-serif font, set against a blue trapezoidal background that tapers to the right.

Facebook Graduate Fellowship Program

Every day Facebook confronts the most complex technical problems. We believe that close relationships with the academic community will enable us to address and solve these challenges at a fundamental level.

As part of our ongoing commitment to academic relationships, we are pleased to announce the Facebook Fellowship program to support graduate students in the 2016-17 and 2017-18 school years.

Research Areas

- Computer Architecture
- Data Mining
- Databases
- Distributed Systems, Networking & Operating Systems
- Economics and Computation
- Human-Computer Interaction & Social Computing
- Machine Learning
- Natural Language Processing
- Programming Languages & Compilers
- Security & Privacy
- Software Engineering

Award Details

- Tuition and fees paid
- \$37,000 yearly grant for two years
- Visit to Facebook HQ to present your research
- Opportunity for paid internship

Important Dates

- October 1, 2015 Application opens
- November 1, 2015 Complete applications with references due
- December 18, 2015 Winners announced

More Information

facebook.com/fellowship
facebookfellowship@fb.com

The Facebook logo, consisting of the word "facebook" in a white, lowercase, sans-serif font.



**A database so scalable
everyone can use it.**

COMMUNITY **and**
ENTERPRISE
Editions

memsql.com/download



PERSISTENT

Don't Refresh. Reimagine!



Join the team that obsessively focuses on the “how” of digital transformation, the ideas of API-centricity, design and data-driven applications for building software-driven businesses.

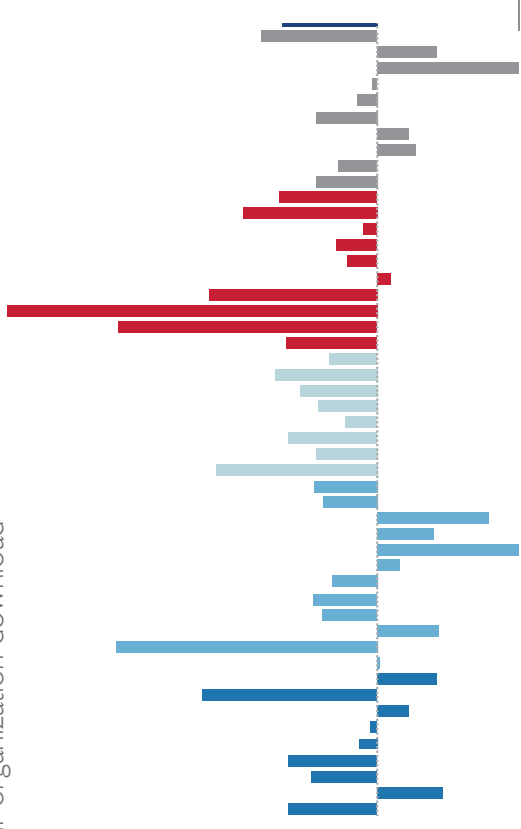
www.persistent.com

Rapid-Fire Business Intelligence for Everyone

- Highest ease of use for developers & business users
- Fastest implementation & report development
- Share & collaborate across your organization download
a free trial at tableau.com/trial



www.tableau.com





Work life that benefits life life. [Learn more](#)



Intel Labs is Looking for Research Scientists!
Submit your resume to: intellabsjobs@intel.com