

2017 IEEE 33rd International Conference on Data Engineering (ICDE 2017)

**San Diego, California, USA
19-22 April 2017**

Pages 809-1616



**IEEE Catalog Number: CFP17026-POD
ISBN: 978-1-5090-6544-8**

**Copyright © 2017 by the Institute of Electrical and Electronics Engineers, Inc
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP17026-POD
ISBN (Print-On-Demand):	978-1-5090-6544-8
ISBN (Online):	978-1-5090-6543-1
ISSN:	1063-6382

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2017 IEEE 33rd International Conference on Data Engineering

ICDE 2017

Table of Contents

Message from the Chairs.....	xxv
Organizing Committee.....	xxvii
Reviewers.....	xxix

Keynotes

Mosaics: Stratosphere, Flink and Beyond	3
<i>Volker Markl</i>	
Leveraging Data and People to Accelerate Data Science	4
<i>Laura Haas</i>	

TKDE Posters

Semantic Bootstrapping: A Theoretical Perspective	7
<i>Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Q. Zhu</i>	
IF-Matching: Towards Accurate Map-Matching with Information Fusion	9
<i>Gang Hu, Jie Shao, Fenglin Liu, Yuan Wang, and Heng Tao Shen</i>	
A Mixed Generative-Discriminative Based Hashing Method	11
<i>Qi Zhang, Yang Wang, Jin Qian, Binbin Deng, and Xuanjing Huang</i>	
SPIRIT: A Tree Kernel-Based Method for Topic Person Interaction Detection (Extended Abstract)	13
<i>Yung-Chun Chang, Chien Chin Chen, and Wen-Lian Hsu</i>	
Efficient Cache-Supported Path Planning on Roads (Extended Abstract)	15
<i>Ying Zhang, Yu-Ling Hsueh, Wang-Chien Lee, and Yi-Hao Jhang</i>	
Personalized Influential Topic Search via Social Network Summarization	17
<i>Jianxin Li, Chengfei Liu, Jeffrey Xu Yu, Yi Chen, Timos Sellis, and J. Shane Culpepper</i>	
Recommendation for Repeat Consumption from User Implicit Feedback	19
<i>Jun Chen, Chaokun Wang, Jianmin Wang, and Philip S. Yu</i>	

PINOCCHIO: Probabilistic Influence-Based Location Selection over Moving Objects	21
<i>Meng Wang, Hui Li, Jiangtao Cui, Ke Deng, Sourav S. Bhowmick, and Zhenhua Dong</i>	
K-Join: Knowledge-Aware Similarity Join	23
<i>Zeyuan Shang, Yaxiao Liu, Guoliang Li, and Jianhua Feng</i>	
Mining Suspicious Tax Evasion Groups in Big Data	25
<i>Feng Tian, Tian Lan, Qinghua Zheng, Kuo-Ming Chao, Nick Godwin, Nazaraf Shah, and Fan Zhang</i>	
Influence Maximization in Trajectory Databases	27
<i>Long Guo, Dongxiang Zhang, Gao Cong, Wei Wu, and Kian-Lee Tan</i>	
A Generic Method for Accelerating LSH-Based Similarity Join Processing (Extended Abstract)	29
<i>Chenyun Yu, Sarana Nutanong, Hangyu Li, Cong Wang, and Xingliang Yuan</i>	
The Moving K Diversified Nearest Neighbor Query	31
<i>Yu Gu, Guanli Liu, Jianzhong Qi, Hongfei Xu, Ge Yu, and Rui Zhang</i>	
The Interaction Between Schema Matching and Record Matching in Data Integration (Extended Abstract)	33
<i>Binbin Gu, Zhixu Li, Xiangliang Zhang, An Liu, Guanfeng Liu, Kai Zheng, Lei Zhao, and Xiaofang Zhou</i>	
Learning Mixtures of Markov Chains from Aggregate Data with Structural Constraints (Extended Abstract)	35
<i>Dixin Luo, Hongteng Xu, Yi Zhen, Bistra Dilkina, Hongyuan Zha, Xiaokang Yang, and Wenjun Zhang</i>	
Patient Flow Prediction via Discriminative Learning of Mutually-Correcting Processes (Extended Abstract)	37
<i>Hongteng Xu, Weichang Wu, Shamim Nemati, and Hongyuan Zha</i>	
Crowdsourced Data Management: A Survey	39
<i>Guoliang Li, Jiannan Wang, Yudian Zheng, and Michael Franklin</i>	
Conflict-Aware Weighted Bipartite b-Matching and Its Application to E-Commerce	41
<i>Cheng Chen, Lan Zheng, Venkatesh Srinivasan, Alex Thomo, Kui Wu, and Anthony Sukow</i>	
Mutually Beneficial Confluent Routing	43
<i>Xinpeng Zhang, Yasuhito Asano, and Masatoshi Yoshikawa</i>	
Exploit Every Bit: Effective Caching for High-Dimensional Nearest Neighbor Search	45
<i>Bo Tang, Man Lung Yiu, and Kien A. Hua</i>	

Bridging Feature Selection and Extraction: Compound Feature Generation (Extended Abstract)	47
<i>Sreevani and C.A. Murthy</i>	
A Novel Cost-Based Model for Data Repairing	49
<i>Shuang Hao, Nan Tang, Guoliang Li, Jian He, Na Ta, and Jianhua Feng</i>	
Fast Memory Efficient Local Outlier Detection in Data Streams (Extended Abstract)	51
<i>Mahsa Salehi, Christopher Leckie, James C. Bezdek, Tharshan Vaithianathan, and Xuyun Zhang</i>	
Finding Causality and Responsibility for Probabilistic Reverse Skyline Query Non-Answers	53
<i>Yunjun Gao, Qing Liu, Gang Chen, Linlin Zhou, and Baihua Zheng</i>	
Bring Order into the Samples: A Novel Scalable Method for Influence Maximization (Extended Abstract)	55
<i>Xiaoyang Wang, Ying Zhang, Wenjie Zhang, Xuemin Lin, and Chen Chen</i>	
Scalable Temporal Latent Space Inference for Link Prediction in Dynamic Social Networks (Extended Abstract)	57
<i>Linhong Zhu, Dong Guo, Junming Yin, Greg Ver Steeg, and Aram Galstyan</i>	
Collective Travel Planning in Spatial Networks	59
<i>Shuo Shang, Lisi Chen, Zhewei Wei, Christian S. Jensen, Ji-Rong Wen, and Panos Kalnis</i>	
Proxies for Shortest Path and Distance Queries	61
<i>Shuai Ma, Kaiyu Feng, Jianxin Li, Haixun Wang, Gao Cong, and Jinpeng Huai</i>	
gMark: Schema-Driven Generation of Graphs and Queries	63
<i>Guillaume Bagan, Angela Bonifati, Radu Ciucanu, George H. L. Fletcher, Aurélien Lemay, and Nicky Advokaat</i>	
SEDEX: Scalable Entity Preserving Data Exchange	65
<i>Yoones A. Sekhavat and Jeffrey Parsons</i>	
Efficient Distributed Density Peaks for Clustering Large Data Sets in MapReduce	67
<i>Yanfeng Zhang, Shimin Cheny, and Ge Yu</i>	
Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage	69
<i>Junbeom Hur, Dongyoung Koo, Youngjoo Shin, and Kyungtae Kang</i>	
Maximizing Acceptance in Rejection-aware Spatial Crowdsourcing	71
<i>Libin Zheng and Lei Chen</i>	

ICDE Short Paper Posters

Mobi-SAGE: A Sparse Additive Generative Model for Mobile App Recommendation	75
<i>Hongzhi Yin, Liang Chen, Weiqing Wang, Xingzhong Du, Quoc Viet Hung Nguyen, and Xiaofang Zhou</i>	
Clustering with Adaptive Manifold Structure Learning	79
<i>Zhao Kang, Chong Peng, and Qiang Cheng</i>	
Mining Precise-Positioning Episode Rules from Event Sequences	83
<i>Xiang Ao, Ping Luo, Jin Wang, Fuzhen Zhuang, and Qing He</i>	
NetClus: A Scalable Framework for Locating Top-K Sites for Placement of Trajectory-Aware Services	87
<i>Shubhadip Mitra, Priya Saraf, Richa Sharma, Arnab Bhattacharya, Sayan Ranuy, and Harsh Bhandari</i>	
Preserving-Ignoring Transformation Based Index for Approximate k Nearest Neighbor Search	91
<i>Gang Hu, Jie Shao, Dongxiang Zhang, Yang Yang, and Heng Tao Shen</i>	
Scalable Top-K Structural Diversity Search	95
<i>Lijun Chang, Chen Zhang, Xuemin Lin, and Lu Qin</i>	
K-Dominant Skyline Join Queries: Extending the Join Paradigm to K-Dominant Skylines	99
<i>Anuradha Awasthi, Arnab Bhattacharya, Sanchit Gupta, and Ujjwal Kumar Singh</i>	
SF-sketch: A Fast, Accurate, and Memory Efficient Data Structure to Store Frequencies of Data Items	103
<i>Tong Yang, Lingtong Liu, Yibo Yan, Muhammad Shahzad, Yulong Shen, Xiaoming Li, Bin Cui, and Gaogang Xie</i>	
Direction-Aware Why-Not Spatial Keyword Top-k Queries	107
<i>Lei Chen, Yafei Li, Jianliang Xu, and Christian S. Jensen</i>	
Searching Time Series with Invariance to Large Amounts of Uniform Scaling	111
<i>Yilin Shen, Yanping Chen, Eamonn Keogh, and Hongxia Jin</i>	
Mining Maximal Cliques on Dynamic Graphs Efficiently by Local Strategies	115
<i>Shengli Sun, Yimo Wang, Weilong Liao, and Wei Wang</i>	
Parallelizing Skip Lists for In-Memory Multi-Core Database Systems	119
<i>Zhongle Xie, Qingchao Cai, H.V. Jagadish, Beng Chin Ooi, and Weng-Fai Wong</i>	
From Raw Footprints to Personal Interests: Bridging the Semantic Gap via Trip Intention Aggregation	123
<i>Long Guo, Dongxiang Zhang, Huayu Wu, Bin Cui, and Kian-Lee Tan</i>	
CrowdFusion: A Crowdsourced Approach on Data Fusion Refinement	127
<i>Yunfan Chen, Lei Chen, and Chen Jason Zhang</i>	

Correcting and Speeding-Up Bounds for Non-Uniform Graph Edit Distance	131
<i>David B. Blumenthal and Johann Gamper</i>	
Distance-Aware Encoding of Numerical Values for Privacy-Preserving Record Linkage	135
<i>Dimitrios Karapiperis, Aris Gkoulalas-Divanis, and Vassilios S. Verykios</i>	
Query Optimizations over Decentralized RDF Graphs	139
<i>Ibrahim Abdelaziz, Essam Mansour, Mourad Ouzzani, Ashraf Aboulnaga, and Panos Kalnis</i>	
LTD-RBM: Robust and Fast Latent Truth Discovery Using Restricted Boltzmann Machines	143
<i>Klaus Broelemann, Thomas Gottron, and Gjergji Kasneci</i>	
A Comparative Study of Log-Structured Merge-Tree-Based Spatial Indexes for Big Data	147
<i>Young-Seok Kim, Taewoo Kim, Michael J. Carey, and Chen Li</i>	
Skew-Resistant Graph Partitioning	151
<i>Angen Zheng, Alexandros Labrinidis, and Christos Faloutsos</i>	
Efficient Multistream Classification Using Direct Density Ratio Estimation	155
<i>Ahsanul Haque, Swarup Chandra, Latifur Khan, Kevin Hamlen, and Charu Aggarwal</i>	
A Distance Measure for the Analysis of Polar Opinion Dynamics in Social Networks	159
<i>Victor Amelkin, Petko Bogdanov, and Ambuj K. Singh</i>	
Cross-Network Clustering and Cluster Ranking for Medical Diagnosis	163
<i>Jingchao Ni, Hongliang Fei, Wei Fan, and Xiang Zhang</i>	
Reverse Query-Aware Locality-Sensitive Hashing for High-Dimensional Furthest Neighbor Search	167
<i>Qiang Huang, Jianlin Feng, and Qiong Fang</i>	
Ontology- and Sentiment-Aware Review Summarization	171
<i>Nhat X.T. Le, Vagelis Hristidis, and Neal Young</i>	
ACTS: An Active Learning Method for Time Series Classification	175
<i>Fengchao Peng, Qiong Luo, and Lionel M. Ni</i>	
FuseM: Query-Centric Data Fusion on Structured Web Markup	179
<i>Ran Yu, Ujwal Gadiraju, Besnik Fetahu, and Stefan Dietze</i>	
Scalable Processing of Massive Uncertain Graph Data: A Simultaneous Processing Approach	183
<i>Zhaonian Zou, Faming Li, Jianzhong Li, and Yingshu Li</i>	
On Edge Classification in Networks with Structure and Content	187
<i>Charu C. Aggarwal, Yao Li, Philip S. Yu, and Yuchen Zhao</i>	

Adaptive State Space Partitioning of Markov Decision Processes for Elastic Resource Management	191
<i>Konstantinos Lolos, Ioannis Konstantinou, Verena Kantere, and Nectarios Koziris</i>	
Sampling and Reconstruction Using Bloom Filters	195
<i>Neha Sengupta, Amitabha Bagchi, Srikanta Bedathur, and Maya Ramanath</i>	
Top-k Frequent Items and Item Frequency Tracking over Sliding Windows of Any Sizes	199
<i>Chunyao Song, Xuanming Liu, and Tingjian Ge</i>	
Joint Gaussian Based Measures for Multiple-Instance Learning	203
<i>Linfei Zhou, Claudia Plant, and Christian Böhm</i>	
Answering Location-Aware Graph Reachability Queries on GeoSocial Data	207
<i>Mohamed Sarwat and Yuhan Sun</i>	
IRanker: Query-Specific Ranking of Reviewed Items	211
<i>Moloud Shahbazi, Matthew Wiley, and Vagelis Hristidis</i>	
Analyzing and Visualizing Scalar Fields on Graphs	215
<i>Yang Zhang, Yusu Wang, and Srinivasan Parthasarathy</i>	
Enterprise Community Detection	219
<i>Jiawei Zhang, Philip S. Yu, and Yuanhua Lv</i>	
Trading Data in Good Faith: Integrating Truthfulness and Privacy Preservation in Data Markets	223
<i>Chaoyue Niu, Zhenzhe Zheng, Fan Wu, Xiaofeng Gao, and Guihai Chen</i>	
Integrative Dynamic Reconfiguration in a Parallel Stream Processing Engine	227
<i>Kasper Grud Skat Madsen, Yongluan Zhou, and Jianneng Cao</i>	
Secure KNN Queries over Encrypted Data: Dimensionality Is Not Always a Curse	231
<i>Xinyu Lei, Alex X. Liu, and Rui Li</i>	

Industry Posters

Improving Predictable Shared-Disk Clusters Performance for Database Clouds	237
<i>Anjan Kumar Amirishetty, Yunrui Li, Tolga Yurek, Mahesh Girkar, Wilson Chan, Graham Ivey, Vsevolod Panteleenko, and Ken Wong</i>	
DeepSD: Supply-Demand Prediction for Online Car-Hailing Services Using Deep Neural Networks	243
<i>Dong Wang, Wei Cao, Jian Li, and Jieping Ye</i>	
Dynamic Statistics Collection in the Teradata Unified Data Architecture	255
<i>Sung Jin Kim, Mohammed Al-Kateb, Paul Sinclair, Alain Crolotte, Chengyang Zhang, and Linda Rose</i>	

Hot or Not? Forecasting Cellular Network Hot Spots Using Sector Performance Indicators	259
<i>Joan Serrà, Ilias Leontiadis, Alexandros Karatzoglou, and Konstantina Papagiannaki</i>	
Joint User-Entity Representation Learning for Event Recommendation in Social Network	271
<i>Lijun Tang and Eric Yi Liu</i>	
TencentBoost: A Gradient Boosting Tree System with Parameter Server	281
<i>Jie Jiang, Jiawei Jiang, Bin Cui, and Ce Zhang</i>	
SSD-Assisted Backup and Recovery for Database Systems	285
<i>Yongseok Son, Jaeyoon Choi, Jekyeom Jeon, Cheolgi Min, Sunggon Kim, Heon Young Yeom, and Hyuck Han</i>	
The Good, the Bad, and the KPIs: How to Combine Performance Metrics to Better Capture Underperforming Sectors in Mobile Networks	297
<i>Ilias Leontiadis, Joan Serrà, Alessandro Finamore, Giorgos Dimopoulos, and Konstantina Papagiannaki</i>	
Too Big to Eat: Boosting Analytics Data Ingestion from Object Stores with Scoop	309
<i>Yosef Moatti, Eran Rom, Raul Gracia-Tinedo, Dalit Naor, Doron Chen, Josep Sampe, Marc Sanchez-Artigas, Pedro Garcia-Lopez, Filip Gluszak, Eric Deschdt, Francesco Pace, Daniele Venzano, and Pietro Michiardi</i>	

Research Track

Session: Graphs

UniWalk: Unidirectional Random Walk Based Scalable SimRank Computation over Large Graph	325
<i>Xiongcai Luo, Jun Gao, Chang Zhou, and Jeffrey Xu Yu</i>	
A Fast Order-Based Approach for Core Maintenance	337
<i>Yikai Zhang, Jeffrey Xu Yu, Ying Zhang, and Lu Qin</i>	
Scalable and Interactive Graph Clustering Algorithm on Multicore CPUs	349
<i>Son T. Mai, Martin Storgaard Dieu, Ira Assent, Jon Jacobsen, Jesper Kristensen, and Mathias Birk</i>	
Fast Computation of Dense Temporal Subgraphs	361
<i>Shuai Ma, Renjun Hu, Luoshu Wang, Xuelian Lin, and Jinpeng Huai</i>	

Session: Keyword Search, Text and Strings

Reverse Keyword-Based Location Search	375
<i>Xike Xie, Xin Lin, Jianliang Xu, and Christian S. Jensen</i>	
Reverse Top-k Geo-Social Keyword Queries in Road Networks	387
<i>Jingwen Zhao, Yunjun Gao, Gang Chen, Christian S. Jensen, Rui Chen, and Deng Cai</i>	
BWT Arrays and Mismatching Trees: A New Way for String Matching with k Mismatches	399
<i>Yangjun Chen and Yujia Wu</i>	
Source-LDA: Enhancing Probabilistic Topic Models Using Prior Knowledge Sources	411
<i>Justin Wood, Patrick Tan, Wei Wang, and Corey Arnold</i>	

Session: Data Mining

Network Backboning with Noisy Data	425
<i>Michele Coscia and Frank M.H. Neffke</i>	
Scalable Informative Rule Mining	437
<i>Guoyao Feng, Lukasz Golab, and Divesh Srivastava</i>	
Streaming k-Means Clustering with Fast Queries	449
<i>Yu Zhang, Kanat Tangwongsan, and Srikanta Tirthapura</i>	
Density Based Clustering over Location Based Services	461
<i>Md Farhadur Rahman, Weimo Liu, Saad Bin Suhaim, Saravanan Thirumuruganathan, Nan Zhang, and Gautam Das</i>	

Session: Query Optimization and Provenance

Provenance-Aware Query Optimization	473
<i>Xing Niu, Raghav Kapoor, Boris Glavic, Dieter Gawlick, Zhen Hua Liu, and Venkatesh Radhakrishnan</i>	
A SQL-Middleware Unifying Why and Why-Not Provenance for First-Order Queries	485
<i>Seokki Lee, Sven Köhler, Bertram Ludäscher, and Boris Glavic</i>	
Extended Characteristic Sets: Graph Indexing for SPARQL Query Optimization	497
<i>Marios Meimaris, George Papastefanatos, Nikos Mamoulis, and Ioannis Anagnostopoulos</i>	
TT-Join: Efficient Set Containment Join	509
<i>Jianye Yang, Wenjie Zhang, Shiyu Yang, Ying Zhang, and Xuemin Lin</i>	

Session: Systems for New Analytics

Scalable Linear Algebra on a Relational Database System	523
<i>Shangyu Luo, Zekai J. Gao, Michael Gubanov, Luis L. Perez, and Christopher Jermaine</i>	
KeystoneML: Optimizing Pipelines for Large-Scale Advanced Analytics	535
<i>Evan R. Sparks, Shivaram Venkataraman, Tomer Kaftan, Michael J. Franklin, and Benjamin Recht</i>	
Parallel SPARQL Query Optimization	547
<i>Buwen Wu, Yongluan Zhou, Hai Jin, and Amol Deshpande</i>	
Efficient Scalable Accurate Regression Queries in In-DBMS Analytics	559
<i>Christos Anagnostopoulos and Peter Triantafillou</i>	
Towards Unified Data and Lifecycle Management for Deep Learning	571
<i>Hui Miao, Ang Li, Larry S. Davis, and Amol Deshpande</i>	

Session: Top-K, KNN and Skyline Querying

Monitoring the Top-m Rank Aggregation of Spatial Objects in Streaming Queries	585
<i>Farhana M. Choudhury, Zhifeng Bao, J. Shane Culpepper, and Timos Sellis</i>	
Answering Top-k Exemplar Trajectory Queries	597
<i>Sheng Wang, Zhifeng Bao, J. Shane Culpepper, Timos Sellis, Mark Sanderson, and Xiaolin Qin</i>	
V-Tree: Efficient kNN Search on Moving Objects with Road-Network Constraints	609
<i>Bilong Shen, Ying Zhao, Guoliang Li, Weimin Zheng, Yue Qin, Bo Yuan, and Yongming Rao</i>	
Sweet KNN: An Efficient KNN on GPU through Reconciliation between Redundancy Removal and Regularity	621
<i>Guoyang Chen, Yufei Ding, and Xipeng Shen</i>	
Secure Skyline Queries on Cloud Platform	633
<i>Jinfei Liu, Juncheng Yang, Li Xiong, and Jian Pei</i>	

Session: New Hardware

Accelerating Multi-Column Selection Predicates in Main-Memory - The Elf Approach	647
<i>David Broneske, Veit Köppen, Gunter Saake, and Martin Schäler</i>	

Revisiting the Design of Data Stream Processing Systems on Multi-Core Processors	659
<i>Shuhao Zhang, Bingsheng He, Daniel Dahlmeier, Amelie Chi Zhou, and Thomas Heinze</i>	
DIDO: Dynamic Pipelines for In-Memory Key-Value Stores on Coupled CPU-GPU Architectures	671
<i>Kai Zhang, Jiayu Hu, Bingsheng He, and Bei Hua</i>	
On Log-Structured Merge for Solid-State Drives	683
<i>Risi Thonangi and Jun Yang</i>	

Session: Security and Encryption

Adaptively Secure Conjunctive Query Processing over Encrypted Data for Cloud Computing	697
<i>Rui Li and Alex X. Liu</i>	
Towards a Unifying Attribute Based Access Control Approach for NoSQL Datastores	709
<i>Pietro Colombo and Elena Ferrari</i>	
Frequency-Hiding Dependency-Preserving Encryption for Outsourced Databases	721
<i>Boxiang Dong and Wendy Wang</i>	
Secure and Efficient Query Processing over Hybrid Clouds	733
<i>Kerim Yasin Oktay, Murat Kantarcioglu, and Sharad Mehrotra</i>	

Session: Similarity Search

Capturing the Moment: Lightweight Similarity Computations	747
<i>Georgios Damaskinos, Rachid Guerraoui, and Rhicheek Patra</i>	
An Efficient Framework for Exact Set Similarity Search Using Tree Structure Indexes	759
<i>Yong Zhang, Xiuxing Li, Jin Wang, Ying Zhang, Chunxiao Xing, and Xiaojie Yuan</i>	
Role Discovery in Graphs Using Global Features: Algorithms, Applications and a Novel Evaluation Strategy	771
<i>Pratik Vinay Gupte, Balaraman Ravindran, and Srinivasan Parthasarathy</i>	
Similarity Search in Graph Databases: A Multi-Layered Indexing Approach	783
<i>Yongjiang Liang and Peixiang Zhao</i>	

Session: Potpourri

Posterior Snapshot Isolation	797
<i>Xuan Zhou, Xin Zhou, Zhengtai Yu, and Kian-Lee Tan</i>	
PrivSuper: A Superset-First Approach to Frequent Itemset Mining under Differential Privacy	809
<i>Ning Wang, Xiaokui Xiao, Yin Yang, Zhenjie Zhang, Yu Gu, and Ge Yu</i>	
Quantifying Differential Privacy under Temporal Correlations	821
<i>Yang Cao, Masatoshi Yoshikawa, Yonghui Xiao, and Li Xiong</i>	
Tracking Matrix Approximation over Distributed Sliding Windows	833
<i>Haida Zhang, Zengfeng Huang, Zhewei Wei, Wenjie Zhang, and Xuemin Lin</i>	

Session: Social Networks

Temporal Influence Blocking: Minimizing the Effect of Misinformation in Social Networks	847
<i>Chonggang Song, Wynne Hsu, and Mong Li Lee</i>	
Complex Event-Participant Planning and Its Incremental Variant	859
<i>Yurong Cheng, Ye Yuan, Lei Chen, Christophe Giraud-Carrier, and Guoren Wang</i>	
Most Influential Community Search over Large Social Networks	871
<i>Jianxin Li, Xinjue Wang, Ke Deng, Xiaochun Yang, Timos Sellis, and Jeffrey Xu Yu</i>	
Boosting Information Spread: An Algorithmic Approach	883
<i>Yishi Lin, Wei Chen, and John C.S. Lui</i>	

Session: Data Cleaning

Cleaning Data with Forbidden Itemsets	897
<i>Joeri Rammelaere, Floris Geerts, and Bart Goethals</i>	
Parallel Progressive Approach to Entity Resolution Using MapReduce	909
<i>Yasser Altowim and Sharad Mehrotra</i>	
A Collective, Probabilistic Approach to Schema Mapping	921
<i>Angelika Kimmig, Alex Memory, Ren´ee J. Miller, and Lise Getoor</i>	
Cleaning Relations Using Knowledge Bases	933
<i>Shuang Hao, Nan Tang, Guoliang Li, and Jian Li</i>	

Session: Learning and Outlier Detection

Time Series Classification by Sequence Learning in All-Subsequence Space	947
<i>Thach Le Nguyen, Severin Gsponer, and Georgiana Ifrim</i>	

Multi-Tactic Distance-Based Outlier Detection	959
<i>Lei Cao, Yizhou Yan, Caitlin Kuhlman, Qingyang Wang, Elke A. Rundensteiner, and Mohamed Eltabakh</i>	
Link Prediction across Aligned Networks with Sparse and Low Rank Matrix Estimation	971
<i>Jiawei Zhang, Jianhui Chen, Shi Zhi, Yi Chang, Philip S. Yu, and Jiawei Han</i>	
LSHiForest: A Generic Framework for Fast Tree Isolation Based Ensemble Anomaly Analysis	983
<i>Xuyun Zhang, Wanchun Dou, Qiang He, Rui Zhou, Christopher Leckie, Ramamohanarao Kotagiri, and Zoran Salcic</i>	

Session: Crowdsourcing and Recommender Systems

Prediction-Based Task Assignment in Spatial Crowdsourcing	997
<i>Peng Cheng, Xiang Lian, Lei Chen, and Cyrus Shahabi</i>	
Trichromatic Online Matching in Real-Time Spatial Crowdsourcing	1009
<i>Tianshu Song, Yongxin Tong, Libin Wang, Jieying She, Bin Yao, Lei Chen, and Ke Xu</i>	
Tuning Crowdsourced Human Computation	1021
<i>Chen Cao, Jiayang Tu, Zheng Liu, Lei Chen, and H.V. Jagadish</i>	
Scalable and Interpretable Product Recommendations via Overlapping Co-Clustering	1033
<i>Reinhard Heckel, Michail Vlachos, Thomas Parnell, and Celestine Duenner</i>	

Session: Distributed Processing

In-Memory Distributed Matrix Computation Processing and Optimization	1047
<i>Yongyang Yu, Mingjie Tang, Walid G. Aref, Qutaibah M. Malluhi, Mostafa M. Abbas, and Mourad Ouzzani</i>	
Fast and Scalable Distributed Set Similarity Joins for Big Data Analytics	1059
<i>Chuitian Rong, Chunbin Lin, Yasin N. Silva, Jianguo Wang, Wei Lu, and Xiaoyong Du</i>	
Fast and Scalable Distributed Boolean Tensor Factorization	1071
<i>Namyong Park, Sejoon Oh, and U. Kang</i>	
Spinner: Scalable Graph Partitioning in the Cloud	1083
<i>Claudio Martella, Dionysios Logothetis, Andreas Loukas, and Georgos Siganos</i>	
Distributed Publish/Subscribe Query Processing on the Spatio-Textual Data Stream	1095
<i>Zhida Chen, Gao Cong, Zhenjie Zhang, Tom Z.J. Fuz, and Lisi Chen</i>	

Industry Track

Session: Predictive Analytics

Predictive Provisioning: Efficiently Anticipating Usage in Azure SQL Database	1111
<i>Lalitha Viswanathan, Bikash Chandra, Willis Lang, Karthik Ramachandra, Jignesh M. Patel, Ajay Kalhan, David J. Dewitt, and Alan Halverson</i>	
Xhare-a-Ride: A Search Optimized Dynamic Ride Sharing System with Approximation Guarantee	1117
<i>Raja Subramaniam Thangaraj, Koyel Mukherjee, Gurulingesh Raravi, Asmita Metrewar, Narendra Annamaneni, and Koushik Chattopadhyay</i>	
Real-Time Novel Event Detection from Social Media	1129
<i>Quanzhi Li, Armineh Nourbakhsh, Sameena Shah, and Xiaomo Liu</i>	
Anomaly Detection in Large Databases Using Behavioral Patterning	1140
<i>Hanna Mazzawi, Gal Dalaly, David Rozenblatz, Liat Ein-Dorx, Matan Niniox, and Ofer Lavi</i>	
The Microsoft Reactive Framework Meets the Internet of Moving Things	1150
<i>Abdeltawab M. Hendawi, Jayant Gupta, Youying Shi, Hossam Fattah, and Mohamed Ali</i>	

Session: New Systems

Twitter Heron: Towards Extensible Streaming Engines	1165
<i>Maosong Fu, Ashvin Agrawal, Avrielia Floratou, Bill Graham, Andrew Jorgensen, Mark Li, Neng Lu, Karthik Ramasamy, Sriram Rao, and Cong Wang</i>	
Feisu: Fast Query Execution over Heterogeneous Data Sources on Large-Scale Clusters	1173
<i>An Qin, Yuan Yuan, Dai Tan, Pengyu Sun, Xiang Zhang, Hao Cao, Rubao Lee, and Xiaodong Zhang</i>	
DistributedLog: A High Performance Replicated Log Service	1183
<i>Sijie Guo, Robin Dhamankar, and Leigh Stewart</i>	
Making Big Data Simple with dashDB Local	1195
<i>Sam Lightstone, Russ Ohanian, Michael Haide, James Cho, Michael Springgay, and Torsten Steinbach</i>	

Session: Optimization and Benchmarks

Optimizing UNION ALL Join Queries in Teradata	1209
<i>Mohammed Al-Kateb, Paul Sinclair, Alain Crolotte, Lu Ma, Grace Au, and Sanjay Nair</i>	

Multi-Query Optimization for Complex Event Processing in SAP ESP	1213
<i>Shuhao Zhang, Hoang Tam Vo, Daniel Dahlmeier, and Bingsheng He</i>	
BigBench V2: The New and Improved BigBench	1225
<i>Ahmad Ghazal, Todor Ivanov, Pekka Kostamaa, Alain Crolotte, Ryan Voong, Mohammed Al-Kateb, Waleed Ghazal, and Roberto V. Zicari</i>	
Providing Enhanced Functionality for Data Store Clients	1237
<i>Arun Iyengar</i>	
Modeling Scalability of Distributed Machine Learning	1249
<i>Alexander Ulanov, Andrey Simanovsky, and Manish Marwah</i>	

Applications Track

Session 1

Automated Personalized Feedback in Introductory Java Programming MOOCs	1259
<i>Victor J. Marin, Tobin Pereira, Srinivas Sridharan, and Carlos R. Rivero</i>	
Vibration Analysis for IoT Enabled Predictive Maintenance	1271
<i>Deokwoo Jung, Zhenjie Zhang, and Marianne Winslett</i>	
Mining Spatio-Temporal Reachable Regions over Massive Trajectory Data	1283
<i>Guojun Wu, Yichen Ding, Yanhua Li, Jie Bao, Yu Zheng, and Jun Luo</i>	
Smart Personalized Routing for Smart Cities	1295
<i>Abdeltawab M. Hendawi, Aqeel Rustum, Amr A. Ahmadain, David Hazel, Ankur Teredesai, Dev Oliver, Mohamed Ali, and John A. Stankovic</i>	

Session 2

Robust Power Line Outage Detection with Unreliable Phasor Measurements	1309
<i>Jose Cordova-Garcia and Xin Wang</i>	
Database System Support for Personalized Recommendation Applications	1320
<i>Mohamed Sarwat, Raha Moraffah, Mohamed F. Mokbel, and James L. Avery</i>	
Efficient Exploration of Telco Big Data with Compression and Decaying	1332
<i>Constantinos Costa, Georgios Chatzimilioudis, Demetrios Zeinalipour-Yazti, and Mohamed F. Mokbel</i>	
Privacy Preserving Anonymity for Periodical SRS Data Publishing	1344
<i>Jie-Teng Wang and Wen-Yang Lin</i>	

Demo Track

Session 1: Cloud, Stream, Query Processing, Provenance

A Learning-Based Service for Cost and Performance Management of Cloud Databases	1361
<i>Ryan Marcus, Sofiya Semenova, and Olga Papaemmanouil</i>	
AdaStorm: Resource Efficient Storm with Adaptive Configuration	1363
<i>Zujian Weng, Qi Guo, Chunkai Wang, Xiaofeng Meng, and Bingsheng He</i>	
AZTEC: A Cloud-based Computational Platform to Integrate Biomedical Resources	1365
<i>Patrick Tan, Yichao Zhou, Xinxin Huang, Giuseppe M. Mazzeo, and Chelsea Ju</i>	
Demonstrating SolveDB: An SQL-Based DBMS for Optimization Applications	1367
<i>Laurynas Šiknys and Torben Bach Pedersen</i>	
HDBExpDetector: Aggregate Sudden-Change Detector over Dynamic Web Databases	1369
<i>Saad Bin Suhaim, Nan Zhang, Gautam Das, and Ali Jaoua</i>	
IS2R: A System for Refining Reverse Top-k Queries	1371
<i>Qing Liu, Yunjun Gao, Linlin Zhou, and Gang Chen</i>	
POLYTICS: Provenance-Based Analytics of Data-Centric Applications	1373
<i>Pierre Bourhis, Daniel Deutch, and Yuval Moskovitch</i>	
Selective In-Place Appends for Real: Reducing Erases on Wear-prone DBMS Storage	1375
<i>Sergey Hardock, Iliia Petrovy, Robert Gottstein, and Alejandro Buchmann</i>	
Understanding the Security Challenges of Oblivious Cloud Storage with Asynchronous Accesses	1377
<i>Cetin Sahin, Aaron Magat, Victor Zakhary, Amr El Abbadi, Huijia (Rachel) Lin, and Stefano Tessaro</i>	
Urd: A Data Summarization Tool for the Acquisition of Meaningful Cardinality Constraints with Probabilistic Intervals	1379
<i>Tania K. Roblot and Sebastian Link</i>	

Session 2: Graph Analytics, Social Networks, Machine Learning

Demonstration of Kite: A Scalable System for Microblogs Data Management	1383
<i>Amr Magdy and Mohamed F. Mokbel</i>	
Adaptive Topic Follow-Up on Twitter	1385
<i>Abdulrahman Alsaudi, Mehdi Sadri, Yasser Altowim, and Sharad Mehrotra</i>	
BEAMS: Bounded Event Detection in Graph Streams	1387
<i>Mohammad Hossein Namaki, Keyvan Sasani, Yinghui Wu, and Tingjian Ge</i>	

GQFast: Fast Graph Exploration with Context-Aware Autocompletion	1389
<i>Chunbin Lin, Jianguo Wang, and Yannis Papakonstantinou</i>	
GscalerCloud: A Web-Based Graph Scaling Service	1391
<i>J.W. Zhang, Anwasha Mal, and Y.C. Tay</i>	
ModelHub: Deep Learning Lifecycle Management	1393
<i>Hui Miao, Ang Li, Larry S. Davis, and Amol Deshpande</i>	
Privacy Cyborg: Towards Protecting the Privacy of Social Media Users	1395
<i>Theodore Georgiou, Amr El Abbadi, and Xifeng Yan</i>	
SocialLens: Searching and Browsing Communities by Content and Interaction	1397
<i>Hongyun Cai, Vincent W. Zheng, Penghe Chen, Fanwei Zhu, Kevin Chen-Chuan Chang, and Zi Huang</i>	
TABOO: Detecting Unstructured Sensitive Information Using Recursive Neural Networks	1399
<i>Jan Neerbeky, Ira Assentz, and Peter Dolog</i>	

Session 3: Applications, Data Visualization, Text Analysis, Data Integration

A Demonstration of TextDB: Declarative and Scalable Text Analytics on Large Data Sets	1403
<i>Zuozhi Wang, Flavio Bayer, Seungjin Lee, Kishore Narendran, Xuxi Pan, Qing Tang, Jimmy Wang, and Chen Li</i>	
A Human-and-Machine Cooperative Framework for Entity Resolution with Quality Guarantees	1405
<i>Zhaoqiang Chen, Qun Chen, and Zhanhuai Li</i>	
A Scalable Data Integration and Analysis Architecture for Sensor Data of Pediatric Asthma	1407
<i>Dimitris Stripelis, José Luis Ambite, Yao-Yi Chiang, Sandrah P. Eckel, and Rima Habre</i>	
DANCE: Data Cleaning with Constraints and Experts	1409
<i>Ahmad Assadi, Tova Milo, and Slava Novgorodov</i>	
DV8: Interactive Analysis of Aviation Data	1411
<i>Behrooz Omidvar-Tehrani, Arnab Nandi, Nicholas Meyer, Dalton Flanagan, and Seth Young</i>	
Hippo in Action: Scalable Indexing of a Billion New York City Taxi Trips and Beyond	1413
<i>Jia Yu, Raha Moraffah, and Mohamed Sarwat</i>	
Landslide Information Service Based on Composition of Physical and Social Sensors	1415
<i>Aibek Musaev and Calton Pu</i>	

MAROON+: A System for Profiling Entities over Time	1417
<i>Furong Li, Mong Li Lee, and Wynne Hsu</i>	
SPATE: Compacting and Exploring Telco Big Data	1419
<i>Constantinos Costa, Georgios Chatzimilioudis, Demetrios Zeinalipour-Yazti, and Mohamed F. Mokbel</i>	
VisFlow: A Visual Database Integration and Workflow Querying System	1421
<i>Xin Mou, Hasan M. Jamil, and Xiaogang Ma</i>	

PhD Symposium Session 1

Keynote: Research with Real Users	1425
<i>Magda Balazinska</i>	
Multiple-Query Optimization of Regular Path Queries	1426
<i>Zahid Abul-Basher</i>	
RELNA: Ranking Attributes in Social Networks to Detect Overlapping Communities Efficiently	1431
<i>Bella Martínez-Seis</i>	

PhD Symposium Session 2

Judgment Analysis Based on Crowdsourced Opinions	1439
<i>Sujoy Chatterjee, Anirban Mukhopadhyay, and Malay Bhattacharyya</i>	
Effects of User Interactions on Online Social Recommender Systems	1444
<i>Anahita Davoudi</i>	

Tutorials

Community Search over Big Graphs: Models, Algorithms, and Opportunities	1451
<i>Xin Huang, Laks V.S. Lakshmanan, and Jianliang Xu</i>	
Bringing Semantics to Spatiotemporal Data Mining: Challenges, Methods, and Applications	1455
<i>Chao Zhang, Quan Yuan, and Jiawei Han</i>	
Web-Scale Blocking, Iterative and Progressive Entity Resolution	1459
<i>Kostas Stefanidis, Vassilis Christophides, and Vasilis Efthymiou</i>	
The Challenges of Global-Scale Data Management	1463
<i>Faisal Nawab, Divyakant Agrawal, and Amr El Abbadi</i>	
Handling Uncertainty in Geo-Spatial Data	1467
<i>Andreas Züfle, Goce Trajcevski, Dieter Pfoser, Matthias Renz, Matthew T. Rice, Timothy Leslie, Paul Delamater, and Tobias Emrich</i>	

Panels

Data Science Education: We're Missing the Boat, Again	1473
<i>Bill Howe, Michael Franklin, Laura Haas, Tim Kraska, and Jeffrey Ullman</i>	
Small Data	1475
<i>Oliver Kennedy, D. Richard Hipp, Stratos Idreos, Amélie Marian, Arnab Nandi, Carmela Troncoso, and Eugene Wu</i>	

HDMM Workshop

Session 1

Fairness in Group Recommendations in the Health Domain	1481
<i>Maria Stratigi, Haridimos Kondylakis, and Kostas Stefanidis</i>	
MASC: Automatic Sleep Stage Classification Based on Brain and Myoelectric Signals	1489
<i>Yuta Suzuki, Makito Sato, Hiroaki Shiokawa, Masashi Yanagisawa, and Hiroyuki Kitagawa</i>	
Survival Topic Models for Predicting Outcomes for Trauma Patients	1497
<i>Yuanyang Zhang, Richard Jiang, and Linda Petzold</i>	

Session 2

Case Study: Classification Algorithms Comparison for the Multi-Label Problem of Chronic Pelvic Pain Diagnosing	1507
<i>Vinicius Oliverio and Omero Bendicto Poli-Neto</i>	
Enabling Real-Time Drug Abuse Detection in Tweets	1510
<i>Nhathai Phan, Soon Ae Chun, Manasi Bhole, and James Geller</i>	
Mother Smoking During Pregnancy and ADHD in Children	1515
<i>Jing (Melody) Yao</i>	

Session 3

Integrated Theory-and Data-Driven Feature Selection in Gene Expression Data Analysis	1525
<i>Vineet K. Raghu, Xiaoyu Ge, Panos K. Chrysanthis, and Panayiotis V. Benos</i>	
A Mortality Study for ICU Patients Using Bursty Medical Events	1533
<i>Luca Bonomi and Xiaoqian Jiang</i>	
Characterizing Organ Donation Awareness from Social Media	1541
<i>Diogo F. Pacheco, Diego Pinheiro, Martin Cadeiras, and Ronaldo Menezes</i>	

DesWeb Workshop

Session 1

Towards Scalable Non-Monotonic Stream Reasoning via Input Dependency Analysis	1553
<i>Thu-Le Pham, Alessandra Mileo, and Muhammad Intizar Ali</i>	
Distance-Based Triple Reordering for SPARQL Query Optimization	1559
<i>Marios Meimaris and George Papastefanatos</i>	
NOUS: Construction and Querying of Dynamic Knowledge Graphs	1563
<i>Sutanay Choudhury, Khushbu Agarwal, Sumit Purohit, Baichuan Zhang, Meg Pirrung, Will Smith, and Mathew Thomas</i>	

Session 2

Ontology-Based Approach for Academic Evaluation System	1569
<i>Siti Aminah, Iis Afriyanti, and Adila Krisnadhi</i>	
PolyFuse: A Large-Scale Hybrid Data Fusion System	1575
<i>Michael Gubanov</i>	
On Recommending Evolution Measures: A Human-Aware Approach	1579
<i>Kostas Stefanidis, Haridimos Kondylakis, and Georgia Troullinou</i>	

Active and HardDB Workshops

HardDB Keynote

The Configurable Cloud - Accelerating Hyperscale Datacenter Services with FPGA	1587
<i>Andre Putnam</i>	

HardDB Papers

Parallel Selectivity Estimation for Optimizing Multidimensional Spatial Join Processing on GPUs	1591
<i>Jianting Zhang, Simin You, and Le Gruenwald</i>	
Are Databases Fit for Hybrid Workloads on GPUs? A Storage Engine's Perspective	1599
<i>Marcus Pinnecke, David Broneske, Gabriel Campero Durand, and Gunter Saake</i>	

Active Invited Talks (Academic)

Data Management Systems on Future Hardware: Challenges and Opportunities	1609
<i>Bingsheng He</i>	
Processing Declarative Queries through Generating Imperative Code in Managed Runtimes	1610
<i>Stratis D. Viglas</i>	
Enabling Effective Utilization of GPUs for Data Management Systems	1612
<i>Xiaodong Zhang</i>	

Active Invited Talks (Industry)

Tradeoffs and Considerations in the Design of Accelerators for Database Applications	1615
<i>Roger Moussalli</i>	
Hardware Acceleration of Database Analytics	1616
<i>Evangelia Sitaridi</i>	

Author Index