









Keynotes

- 1  **Running with Scissors: Fast Queries on Just-in-Time Databases**
(*Anastasia Ailamaki*)
 - 2  **Transforming Big Data into Smart Data: Deriving Value via Harnessing Volume, Variety, and Velocity Using Semantic Techniques and Technologies**
(*Amit Sheth*)
-

Research Session 1: Clustering




- 3  **Incremental Cluster Evolution Tracking from Highly Dynamic Network Data**
(*Pei Lee, Laks V.S. Lakshmanan, Evangelos E. Milios*)
 - 15  **Finding Common Ground Among Experts' Opinions on Data Clustering: With Applications in Malware Analysis**
(*Guanhua Yan*)
 - 28  **Towards Effective and Efficient Mining of Arbitrary Shaped Clusters**
(*Hao Huang, Yunjun Gao, Kevin Chiew, Lei Chen, Qinqing He*)
-

Research Session 2: Distributed Processing




- 40  **R-Store: A Scalable Distributed System for Supporting Real-Time Analytics**
(*Feng Li, M. Tamer Özsu, Gang Chen, Beng Chin Ooi*)
- 52  **Blazes: Coordination Analysis for Distributed Programs**
(*Peter Alvaro, Neil Conway, Joseph M. Hellerstein, David Maier*)
- 64  **Query Optimization of Distributed Pattern Matching**
(*Jiewen Huang, Kartik Venkatraman, Daniel J. Abadi*)






Research Session 3: Data Mining I – Outliers and Time Series

- 76  **Scalable Distance-Based Outlier Detection Over High-Volume Data Streams**
(Lei Cao, Di Yang, Qingyang Wang, Yanwei Yu, Jiayuan Wang, Elke A. Rundensteiner)
- 88  **Discriminative Features for Identifying and Interpreting Outliers**
(Xuan Hong Dang, Ira Assent, Raymond T. Ng, Arthur Zimek, Erich Schubert)
- 100  **Memory-Efficient Centroid Decomposition for Long Time Series**
(Mourad Khayati, Michael Böhlen, Johann Gamper)
-

Research Session 4: Pareto Optimization


- 112  **Incremental Discovery of Prominent Situational Facts**
(Afroza Sultana, Naeemul Hassan, Chengkai Li, Jun Yang, Cong Yu)
- 124  **Continuous Fragmented Skylines Over Distributed Streams**
(Odysseas Papapetrou, Minos Garofalakis)
- 136  **Stochastic Skyline Route Planning Under Time-Varying Uncertainty**
(Bin Yang, Chenjuan Guo, Christian S. Jensen, Manohar Kaul, Shuo Shang)
-

Research Session 5: Keyword Search I – Spatial




- 148  **Scalable Top-k Spatio-Temporal Term Querying**
(Anders Skovsgaard, Darius Sidlauskas, Christian S. Jensen)
- 160  **Nearest Keyword Set Search in Multi-Dimensional Datasets**
(Vishwakarma Singh, Ambuj K. Singh)
- 172  **Mercury: A Memory-Constrained Spatio-Temporal Real-Time Search on Microblogs**
(Amr Magdy, Mohamed F. Mokbel, Sameh Elnikety, Suman Nath, Yuxiong He)






Research Session 6: Graphs I - Fundamental Algorithms

- 184  **Answering Graph Pattern Queries Using Views**
(Wenfei Fan, Xin Wang, Yinghui Wu)
- 196  **Efficient Top-k Closeness Centrality Search**
(Paul W. Olsen Jr., Alan G. Labouseur, Jeong-Hyon Hwang)
- 208  **Contract & Expand: I/O Efficient SCCs Computing**
(Zhiwei Zhang, Lu Qin, Jeffrey Xu Yu)
-

Research Session 7: Schema Matching and Cleaning




- 220  **Pay-as-You-Go Reconciliation in Schema Matching Networks**
(Quoc Viet Hung Nguyen, Thanh Tam Nguyen, Zoltán Miklós, Karl Aberer, Avigdor Gal, Matthias Weidlich)
- 232  **Mapping and Cleaning**
(Floris Geerts, Giansalvatore Mecca, Paolo Papotti, Donatello Santoro)
- 244  **Continuous Data Cleaning**
(Maksims Volkovs, Fei Chiang, Jaroslaw Szlichta, Renée J. Miller)
-

Research Session 8: Keyword Search II - Privacy and Social Data




- 256  **On Masking Topical Intent in Keyword Search**
(Peng Wang, Chinya V. Ravishankar)
- 268  **Keyword-Based Correlated Network Computation Over Large Social Media**
(Jianxin Li, Chengfei Liu, Md. Saiful Islam)
- 280  **Head, Modifier, and Constraint Detection in Short Texts**
(Zhongyuan Wang, Haixun Wang, Zhirui Hu)






Research Session 9: Graphs II - Ranks and Communities

- 292  **LinkSCAN^{*}: Overlapping Community Detection Using the Link-Space Transformation**
(Sungsu Lim, Seungwoo Ryu, Sejeong Kwon, Kyomin Jung, Jae-Gil Lee)
- 304  **Fast Incremental SimRank on Link-Evolving Graphs**
(Weiren Yu, Xuemin Lin, Wenjie Zhang)
- 316  **Fast Top-K Path-Based Relevance Query on Massive Graphs**
(Samamon Khemmarat, Lixin Gao)
-

Research Session 10: Strings and Texts




- 328  **Efficient Instant-Fuzzy Search with Proximity Ranking**
(Inci Cetindil, Jamshid Esmaelnezhad, Taewoo Kim, Chen Li)
- 340  **MassJoin: A MapReduce-Based Method for Scalable String Similarity Joins**
(Dong Deng, Guoliang Li, Shuang Hao, Jiannan Wang, Jianhua Feng)
- 352  **In-RDBMS Inverted Indexes Revisited**
(Ian Rae, Alan Halverson, Jeffrey F. Naughton)
-

Research Session 11: Temporal and Event Data




- 364  **Adaptive Parallel Compressed Event Matching**
(Mohammad Sadoghi, Hans-Arno Jacobsen)
- 376  **Matching Heterogeneous Events with Patterns**
(Xiaochen Zhu, Shaoxu Song, Jianmin Wang, Philip S. Yu, Jianguang Sun)
- 388  **Leveraging Metadata for Identifying Local, Robust Multi-Variate Temporal (RMT) Features**
(Xiaolan Wang, K. Selçuk Candan, Maria Luisa Sapino)






Research Session 12: Personalized Data Management

- 400  **Personalized Query Suggestion with Diversity Awareness**
(Di Jiang, Kenneth Wai-Ting Leung, Jan Vosecky, Wilfred Ng)
- 412  **Exploiting Group Recommendation Functions for Flexible Preferences**
(Senjuti Basu Roy, Saravanan Thirumuruganathan, Sihem Amer-Yahia, Gautam Das, Cong Yu)
- 424  **PAQO: Preference-Aware Query Optimization for Decentralized Database Systems**
(Nicholas L. Farnan, Adam J. Lee, Panos K. Chrysanthis, Ting Yu)
-

Research Session 13: Data Mining II – Pattern Discovery

- 436  **Automatic Generation of Question Answer Pairs from Noisy Case Logs**
(Jitendra Ajmera, Sachindra Joshi, Ashish Verma, Amol Mittal)
- 448  **Complete Discovery of High-Quality Patterns in Large Numerical Tensors**
(Loïc Cerf, Wagner Meira Jr.)
- 460  **Ranking Item Features by Mining Online User-Item Interactions**
(Sofiane Abbar, Habibur Rahman, Saravanan Thirumuruganathan, Carlos Castillo, Gautam Das)
-

Research Session 14: Data Warehousing




- 472  **Distributed and Interactive Cube Exploration**
(Niranjan Kamat, Prasanth Jayachandran, Karthik Tunga, Arnab Nandi)
- 484  **A Tunable Compression Framework for Bitmap Indices**
(Gheorghii Guzun, Guadalupe Canahuate, David Chiu, Jason Sawin)
- 496  **Pagrol: PARallel GRaph OLap Over Large-Scale Attributed Graphs**
(Zhengkui Wang, Qi Fan, Huiju Wang, Kian-Lee Tan, Divyakant Agrawal, Amr El Abbadi)



Research Session 15: Query Optimization





- 508  **Waste Not... Efficient Co-Processing of Relational Data**
(Holger Pirk, Stefan Manegold, Martin Kersten)
- 520  **History-Aware Query Optimization with Materialized Intermediate Views**
(Luis L. Perez, Christopher M. Jermaine)
- 532  **Decorrelation of User Defined Function Invocations in Queries**
(Varun Simhadri, Karthik Ramachandra, Arun Chaitanya, Ravindra Guravannavar, S. Sudarshan)
-

Research Session 16: Graphs III - Distributed Processing





- 544  **GLog: A High Level Graph Analysis System Using MapReduce**
(Jun Gao, Jiashuai Zhou, Chang Zhou, Jeffrey Xu Yu)
- 556  **Continuous Pattern Detection Over Billion-Edge Graph Using Distributed Framework**
(Jun Gao, Chang Zhou, Jiashuai Zhou, Jeffrey Xu Yu)
- 568  **How to Partition a Billion-Node Graph**
(Lu Wang, Yanghua Xiao, Bin Shao, Haixun Wang)



Research Session 17: Main-Memory Databases





- 580  **Exploiting Hardware Transactional Memory in Main-Memory Databases**
(*Viktor Leis, Alfons Kemper, Thomas Neumann*)
- 592  **Locality-Sensitive Operators for Parallel Main-Memory Database Clusters**
(*Wolf Rödiger, Tobias Mühlbauer, Philipp Unterbrunner, Angelika Reiser, Alfons Kemper, Thomas Neumann*)
- 604  **Rethinking Main Memory OLTP Recovery**
(*Nirmesh Malviya, Ariel Weisberg, Samuel Madden, Michael Stonebraker*)
- 616  **Optimal Hierarchical Layouts for Cache-Oblivious Search Trees**
(*Peter Lindstrom, Deepak Rajan*)
-

Research Session 18: Privacy and Security



- 628  **Private Search on Key-Value Stores with Hierarchical Indexes**
(*Haibo Hu, Jianliang Xu, Xizhong Xu, Kexin Pei, Byron Choi, Shuigeng Zhou*)
- 640  **Practical k Nearest Neighbor Queries with Location Privacy**
(*Xun Yi, Russell Paulet, Elisa Bertino, Vijay Varadharajan*)
- 652  **Generating Private Synthetic Databases for Untrusted System Evaluation**
(*Wentian Lu, Gerome Miklau, Vani Gupta*)
- 664  **Secure k -Nearest Neighbor Query Over Encrypted Data in Outsourced Environments**
(*Yousef Elmehdwi, Bharath K. Samanthula, Wei Jiang*)



Research Session 19: Transaction Management

- 676  **Omid: Lock-Free Transactional Support for Distributed Data Stores**
(*Daniel Gómez Ferro, Flavio Junqueira, Ivan Kelly, Benjamin Reed, Maysam Yabandeh*)
- 688  **ATraPos: Adaptive Transaction Processing on Hardware Islands**
(*Danica Porobic, Erietta Liarou, Pinar Tözün, Anastasia Ailamaki*)
- 700  **Scalable Serializable Snapshot Isolation for Multicore Systems**
(*Hyuck Han, SeongJae Park, Hyungsoo Jung, Alan Fekete, Uwe Röhm, Heon Y. Yeom*)
- 712  **Automatic Entity-Grouping for OLTP Workloads**
(*Bin Liu, Junichi Tatemura, Oliver Po, Wang-Pin Hsiung, Hakan Hacigümüş*)
-

Research Session 20: Graphs IV – Random Walk




- 724  **Evaluating Multi-Way Joins Over Discounted Hitting Time**
(*Wangda Zhang, Reynold Cheng, Ben Kao*)
- 736  **Random-Walk Domination in Large Graphs**
(*Rong-Hua Li, Jeffrey Xu Yu, Xin Huang, Hong Cheng*)
-

Research Session 21: Multidimensional Search

- 748  **Top- k Preferences in High Dimensions**
(*Albert Yu, Pankaj K. Agarwal, Jun Yang*)
- 760  **SLICE: Reviving Regions-Based Pruning for Reverse k Nearest Neighbors Queries**
(*Shiyu Yang, Muhammad Aamir Cheema, Xuemin Lin, Ying Zhang*)
- 772  **Geometry Approach for k -Regret Query**
(*Peng Peng, Raymond Chi-Wing Wong*)



Research Session 22: Similarity Joins

- 784  **L2AP: Fast Cosine Similarity Search with Prefix L-2 Norm Bounds**
(David C. Anastasiu, George Karypis)
- 796  **PHiDJ: Parallel Similarity Self-Join for High-Dimensional Vector Data with MapReduce**
(Sergej Fries, Brigitte Boden, Grzegorz Stepień, Thomas Seidl)
- 808  **MELODY-JOIN: Efficient Earth Mover's Distance Similarity Joins Using MapReduce**
(Jin Huang, Rui Zhang, Rajkumar Buyya, Jian Chen)
-

Research Session 23: Subgraph Mining and Matching

- 820  **Top-K Interesting Subgraph Discovery in Information Networks**
(Manish Gupta, Jing Gao, Xifeng Yan, Hasan Cam, Jiawei Han)
- 832  **Cloud Service Placement via Subgraph Matching**
(Bo Zong, Ramya Raghavendra, Mudhakar Srivatsa, Xifeng Yan, Ambuj K. Singh, Kang-Won Lee)
- 844  **Large-Scale Frequent Subgraph Mining in MapReduce**
(Wenqing Lin, Xiaokui Xiao, Gabriel Ghinita)
-

Research Session 24: Social Contents




- 856  **We Can Learn Your #Hashtags: Connecting Tweets to Explicit Topics**
(Wei Feng, Jianyong Wang)
- 868  **Interactive Hierarchical Tag Clouds for Summarizing Spatiotemporal Social Contents**
(Wei Kang, Anthony K.H. Tung, Feng Zhao, Xinyu Li)
- 880  **Effective Location Identification from Microblogs**
(Guoliang Li, Jun Hu, Jianhua Feng, Kian-Lee Tan)



Research Session 25: Uncertain and Probabilistic Data

- 892  **Efficient and Accurate Query Evaluation on Uncertain Graphs via Recursive Stratified Sampling**
(Rong-Hua Li, Jeffrey Xu Yu, Rui Mao, Tan Jin)
- 904  **Subgraph Pattern Matching Over Uncertain Graphs with Identity Linkage Uncertainty**
(Walaa Eldin Moustafa, Angelika Kimmig, Amol Deshpande, Lise Getoor)
- 916  **User-Driven Refinement of Imprecise Queries**
(Bahar Qarabaqi, Mirek Riedewald)
-

Research Session 26: XML and Tree Data




- 928  **A General Algorithm for Subtree Similarity-Search**
(Sara Cohen, Nerya Or)
- 940  **Breaking out of the MisMatch Trap**
(Yong Zeng, Zhifeng Bao, Tok Wang Ling, H.V. Jagadish, Guoliang Li)
- 952  **XQuery Streaming by Forest Transducers**
(Shizuya Hakuta, Sebastian Maneth, Keisuke Nakano, Hideya Iwasaki)
-

Research Session 27: Crowdsourcing




- 964  **Crowd-Powered Find Algorithms**
(Anish Das Sarma, Aditya Parameswaran, Hector Garcia-Molina, Alon Halevy)
- 976  **A Hybrid Machine-Crowdsourcing System for Matching Web Tables**
(Ju Fan, Meiyu Lu, Beng Chin Ooi, Wang-Chiew Tan, Meihui Zhang)
- 988  **Combining Information Extraction and Human Computing for Crowdsourced Knowledge Acquisition**
(Sarath Kumar Kondreddi, Peter Triantafillou, Gerhard Weikum)



Research Session 28: Spatial and Location Data

- 1000  **OCTOPUS: Efficient Query Execution on Dynamic Mesh Datasets**
(Farhan Tauheed, Thomas Heinis, Felix Schürmann, Henry Markram, Anastasia Ailamaki)
- 1012  **An Efficient Sampling Method for Characterizing Points of Interests on Maps**
(Pinghui Wang, Wenbo He, Xue Liu)
- 1024  **Declarative Cartography: In-Database Map Generalization of Geospatial Datasets**
(Pimin Konstantin Kefaloukos, Marcos Vaz Salles, Martin Zachariasen)
-

Research Session 29: Data Flow and Profiling




- 1036  **Detecting Unique Column Combinations on Dynamic Data**
(Ziawasch Abedjan, Jorge-Arnulfo Quiané-Ruiz, Felix Naumann)
- 1048  **Modeling Data for Business Processes**
(Yutian Sun, Jianwen Su, Budan Wu, Jian Yang)
- 1060  **Engine Independence for Logical Analytic Flows**
(Petar Jovanovic, Alkis Simitsis, Kevin Wilkinson)







Industry Session 1

- 1072  **SILVERBACK: Scalable Association Mining for Temporal Data in Columnar Probabilistic Databases**
(Yusheng Xie, Diana Palsetia, Goce Trajcevski, Ankit Agrawal, Alok Choudhary)
- 1084  **DBDesigner: A Customizable Physical Design Tool for Vertica Analytic Database**
(Ramakrishna Varadarajan, Vivek Bharathan, Ariel Cary, Jaimin Dave, Sreenath Bodagala)
- 1096  **Region Sampling and Estimation of GeoSocial Data with Dynamic Range Calibration**
(Yanhua Li, Moritz Steiner, Jie Bao, Limin Wang, Ting Zhu)
-

Industry Session 2

- 1108  **The Vertica Query Optimizer: The Case for Specialized Query Optimizers**
(Nga Tran, Andrew Lamb, Lakshmikant Shrinivas, Sreenath Bodagala, Jaimin Dave)
- 1120  **Near Neighbor Join**
(Herald Kllapi, Boulos Harb, Cong Yu)
- 1132  **Efficient Support of XQuery Full Text in SQL/XML Enabled RDBMS**
(Zhen Hua Liu, Ying Lu, Hui Joe Chang)
-

Industry Session 3

- 1144  **CrowdPlanner: A Crowd-Based Route Recommendation System**
(Han Su, Kai Zheng, Jiamin Huang, Hoyoung Jeung, Lei Chen, Xiaofang Zhou)
- 1156  **Exploration of the Effect of Category Match Score in Search Advertising**
(Youngchul Cha, Junghoo Cho, Jian Yuan, Tak Yan)
- 1162  **CaSSanDra: An SSD Boosted Key-Value Store**
(Prashanth Menon, Tilmann Rabl, Mohammad Sadoghi, Hans-Arno Jacobsen)
- 1168  **Stock Trade Volume Prediction with Yahoo Finance User Browsing Behavior**
(Ilaria Bordino, Nicolas Kourtellis, Nikolay Laptev, Youssef Billawala)












Demo Session 1

- 1174  **C-DMr: Crowd-Powered Decision Maker for Real World Knapsack Problems**
(*Leihao Xia, Caleb Chen Cao, Lei Chen, Zhao Chen*)
- 1178  **A Crowd-Based Route Recommendation System — CrowdPlanner**
(*Han Su, Kai Zheng, Jiamin Huang, Tianyu Liu, Haozhou Wang, Xiaofang Zhou*)
- 1182  **CrowdCleaner: Data Cleaning for Multi-Version Data on the Web via Crowdsourcing**
(*Yongxin Tong, Caleb Chen Cao, Chen Jason Zhang, Yatao Li, Lei Chen*)
- 1186  **iTag: Incentive-Based Tagging**
(*Siyu Lei, Xuan S. Yang, Luyi Mo, Silviu Maniu, Reynold Cheng*)
- 1190  **Parallel SECONDO: A Practical System for Large-Scale Processing of Moving Objects**
(*Jiamin Lu, Ralf Hartmut Güting*)
- 1194  **Constructing Indoor Navigation Systems from Digital Building Information**
(*Mikkel Boysen, Christian de Haas, Hua Lu, Xike Xie, Aiste Pilvinyte*)
- 1198  **Profiling and Mining RDF Data with ProLOD++**
(*Ziawasch Abedjan, Toni Gruetze, Anja Jentzsch, Felix Naumann*)
- 1202  **Kondenser: Exploration and Visualization of Archived Social Media**
(*Omar Alonso, Kartikay Khandelwal*)
- 1206  **Trendspedia: An Internet Observatory for Analyzing and Visualizing the Evolving Web**
(*Wei Kang, Anthony K.H. Tung, Wei Chen, Xinyu Li, Qiyue Song, Chao Zhang, Feng Zhao, Xiajuan Zhou*)












Demo Session 2

- 1210  **AQUAS: A Quality-Aware Scheduler for NoSQL Data Stores**
(Chen Xu, Fan Xia, Mohamed A. Sharaf, Minqi Zhou, Aoying Zhou)
- 1214  **Outsourcing Multi-Version Key-Value Stores with Verifiable Data Freshness**
(Yuzhe Tang, Ting Wang, Xin Hu, Reiner Sailer, Ling Liu, Peter Pietzuch)
- 1218  **IQ-METER — An Evaluation Tool for Data-Transformation Systems**
(Giansalvatore Mecca, Paolo Papotti, Donatello Santoro)
- 1222  **RULEMINER: Data Quality Rules Discovery**
(Xu Chu, Ihab F. Ilyas, Paolo Papotti, Yin Ye)
- 1226  **iCoDA: Interactive and Exploratory Data Completeness Analysis**
(Ruilin Liu, Guan Wang, Wendy Hui Wang, Flip Korn)
- 1230  **VoidWiz: Resolving Incompleteness Using Network Effects**
(Christina Christodoulakis, Christos Faloutsos, Renée J. Miller)
- 1234  **ADaPT: Automatic Data Personalization Based on Contextual Preferences**
(Antonio Miele, Elisa Quintarelli, Emanuele Rabosio, Letizia Tanca)
- 1238  **Mars: Real-Time Spatio-Temporal Queries on Microblogs**
(Amr Magdy, Ahmed M. Aly, Mohamed F. Mokbel, Sameh Elnikety, Yuxiong He, Suman Nath)
- 1242  **Pigeon: A Spatial MapReduce Language**
(Ahmed Eldawy, Mohamed F. Mokbel)
- 1246  **A Demonstration of MNTG — A Web-Based Road Network Traffic Generator**
(Mohamed F. Mokbel, Louai Alarabi, Jie Bao, Ahmed Eldawy, Amr Magdy, Mohamed Sarwat, Ethan Waytas, Steven Yackel)



Demo Session 3

- 1250  **GOBE: Querying Knowledge Graphs by Example Entity Tuples**
(Nandish Jayaram, Mahesh Gupta, Arijit Khan, Chengkai Li, Xifeng Yan, Ramez Elmasri)
- 1254  **KnowLife: A Knowledge Graph for Health and Life Sciences**
(Patrick Ernst, Cynthia Meng, Amy Siu, Gerhard Weikum)
- 1258  **Text and Structured Data Fusion in Data Tamer at Scale**
(Michael Gubanov, Michael Stonebraker, Daniel Bruckner)
- 1262  **dbTouch in Action — Database Kernels for Touch-Based Data Exploration**
(Erietta Liarou, Stratos Idreos)
- 1266  **SAGE: A Logical and Physical Design Tool for Entity-Group Based New SQL Systems**
(Bin Liu, Wang-Pin Hsiung, Junichi Tatemura, Hakan Hacigümüş)
- 1270  **A Tool for Internet-Scale Cardinality Estimation of XPath Queries Over Distributed Semistructured Data**
(Vasil Slavov, Anas Katib, Praveen Rao)
- 1274  **HOPE: Iterative and Interactive Database Partitioning for OLTP Workloads**
(Yu Cao, Xiaoyan Guo, Baoyao Zhou, Stephen Todd)
- 1278  **Devel-Op: An Optimizer Development Environment**
(Zhibo Peng, Mitch Cherniack, Olga Papaemmanouil)
- 1282  **Guaranteed Authenticity and Integrity of Data from Untrusted Servers**
(Rohit Jain, Sunil Prabhakar)



Tutorials

- 1286  **Linked Data Query Processing**
(Olaf Hartig, M. Tamer Özsu)
- 1290  **Data Stream Warehousing**
(Lukasz Golab, Theodore Johnson)
- 1294  **Data Quality: The Other Face of Big Data**
(Barna Saha, Divesh Srivastava)
- 1298  **Just-in-Time Compilation for SQL Query Processing**
(Stratis D. Viglas)
- 1302  **Managing Uncertainty in Spatial and Spatio-Temporal Data**
(Reynold Cheng, Tobias Emrich, Hans-Peter Kriegel, Nikos Mamoulis, Matthias Renz, Goce Trajcevski, Andreas Züfle)
- 1306  **Distributed Execution of Continuous Queries**
(Rajeev Gupta, Krithi Ramamritham)
-

Panel

- 1310  **Main-Memory Database Systems**
(Alfons Kemper, Thomas Neumann)