

2013 IEEE 29th International Conference on Data Engineering (ICDE 2013)

**Brisbane, Australia
8 - 12 April 2013**

Pages 1-696



**IEEE Catalog Number: CFP13026-PRT
ISBN: 978-1-4673-4909-3**

Keynotes

Hardware Killed the Software Star

(Gustavo Alonso)

Recent Progress Towards an Ecosystem of Structured Data on the Web

(Nitin Gupta, Alon Y. Halevy, Boulos Harb, Heidi Lam, Hongrae Lee, Jayant Madhavan, Fei Wu, Cong Yu)

Re-Thinking the Performance of Information Processing Systems

(Vishal Sikka)

Research Session 1: Main Memory Databases

CPU and Cache Efficient Management of Memory-Resident Databases

(Holger Pirk, Florian Funke, Martin Grund, Thomas Neumann, Ulf Leser, Stefan Manegold, Alfons Kemper, Martin Kersten)

Identifying Hot and Cold Data in Main-Memory Databases

(Justin J. Levandoski, Per-Åke Larson, Radu Stoica)

The Adaptive Radix Tree: ARTful Indexing for Main-Memory Databases

(Viktor Leis, Alfons Kemper, Thomas Neumann)

Research Session 2: MapReduce Algorithms

Finding Connected Components in Map-Reduce in Logarithmic Rounds

(Vibhor Rastogi, Ashwin Machanavajjhala, Laukik Chitnis, Anish Das Sarma)

Enumerating Subgraph Instances Using Map-Reduce

(Foto N. Afrati, Dimitris Fotakis, Jeffrey D. Ullman)

Scalable Maximum Clique Computation Using MapReduce

(Jingen Xiang, Cong Guo, Ashraf Aboulhaga)

Research Session 3: Time Travel in Database

Ficklebase: Looking into the Future to Erase the Past

(Sumeet Bajaj, Radu Sion)

Time Travel in a Scientific Array Database

(Emad Soroush, Magdalena Balazinska)

Time Travel in Column Stores

(Martin Kaufmann, Amin A. Manjili, Stefan Hildenbrand, Donald Kossmann, Andreas Tonder)

Research Session 4: Top-k Query in Uncertain Data

Top-k Query Processing in Probabilistic Databases with Non-Materialized Views

(Maximilian Dylla, Iris Miliaraki, Martin Theobald)

Cleaning Uncertain Data for Top-k Queries

(Luyi Mo, Reynold Cheng, Xiang Li, David W. Cheung, Xuan S. Yang)

Top-K Oracle: A New Way to Present Top-K Tuples for Uncertain Data

(Chunyao Song, Zheng Li, Tingjian Ge)

Research Session 5: Uncertainty in Spatial Data

Voronoi-Based Nearest Neighbor Search for Multi-Dimensional Uncertain Databases

(Peiwu Zhang, Reynold Cheng, Nikos Mamoulis, Matthias Renz, Andreas Züfle, Yu Tang, Tobias Emrich)

Interval Reverse Nearest Neighbor Queries on Uncertain Data with Markov Correlations

(Chuanfei Xu, Yu Gu, Lei Chen, Jianzhong Qiao, Ge Yu)

Efficient Tracking and Querying for Coordinated Uncertain Mobile Objects

(Nicholas D. Larusso, Ambuj Singh)

Research Session 6: Data Extraction

Attribute Extraction and Scoring: A Probabilistic Approach
(Taesung Lee, Zhongyuan Wang, Haixun Wang, Seung-won Hwang)

TYPifier: Inferring the Type Semantics of Structured Data
(Yongtao Ma, Thanh Tran, Veli Bicer)

SUSIE: Search Using Services and Information Extraction
(Nicoleta Preda, Fabian Suchanek, Wenjun Yuan, Gerhard Weikum)

Research Session 7: Trajectory Databases

Towards Efficient Search for Activity Trajectories
(Kai Zheng, Shuo Shang, Nicholas Jing Yuan, Yi Yang)

On Discovery of Gathering Patterns from Trajectories
(Kai Zheng, Yu Zheng, Nicholas Jing Yuan, Shuo Shang)

Destination Prediction by Sub-Trajectory Synthesis and Privacy Protection Against Such Prediction
(Andy Yuan Xue, Rui Zhang, Yu Zheng, Xing Xie, Jin Huang, Zhenghua Xu)

Research Session 8: Social Networks

Scalable and Parallelizable Processing of Influence Maximization for Large-Scale Social Networks
(Jinha Kim, Seung-Keol Kim, Hwanjo Yu)

Socialite: Datalog Extensions for Efficient Social Network Analysis
(Jiwon Seo, Stephen Guo, Monica S. Lam)

LinkProbe: Probabilistic Inference on Large-Scale Social Networks
(Haiquan Chen, Wei-Shinn Ku, Haixun Wang, Liang Tang, Min-Te Sun)

Research Session 9: Indexing Structures

The Bw-Tree: A B-Tree for New Hardware Platforms

(Justin J. Levandoski, David B. Lomet, Sudipta Sengupta)

Secure and Efficient Range Queries on Outsourced Databases Using \hat{R} -Trees

(Peng Wang, Chinya V. Ravishankar)

An Efficient and Compact Indexing Scheme for Large-Scale Data Store

(Peng Lu, Sai Wu, Lidan Shou, Kian-Lee Tan)

Research Session 10: Main Memory Query Processing

Recycling in Pipelined Query Evaluation

(Fabian Nagel, Peter Boncz, Stratis D. Viglas)

Efficient Many-Core Query Execution in Main Memory Column-Stores

(Jonathan Dees, Peter Sanders)

Main-Memory Hash Joins on Multi-Core CPUs: Tuning to the Underlying Hardware

(Cagri Balkesen, Jens Teubner, Gustavo Alonso, M. Tamer Özsu)

Research Session 11: Data Mining I

Coupled Clustering Ensemble: Incorporating Coupling Relationships Both Between Base Clusterings and Objects

(Can Wang, Zhong She, Longbing Cao)

Focused Matrix Factorization for Audience Selection in Display Advertising

(Bhargav Kanagal, Amr Ahmed, Sandeep Pandey, Vanja Josifovski, Lluís Garcia-Pueyo, Jeff Yuan)

Graph Stream Classification Using Labeled and Unlabeled Graphs

(Shirui Pan, Xingquan Zhu, Chengqi Zhang, Philip S. Yu)

Research Session 12: Moving Objects

T-Share: A Large-Scale Dynamic Taxi Ridesharing Service

(Shuo Ma, Yu Zheng, Ouri Wolfson)

Efficient Notification of Meeting Points for Moving Groups via Independent Safe Regions

(Jing Li, Man Lung Yiu, Nikos Mamoulis)

Efficient Distance-Aware Query Evaluation on Indoor Moving Objects

(Xike Xie, Hua Lu, Torben Bach Pedersen)

Research Session 13: Data Cleaning

HANDS: A Heuristically Arranged Non-Backup In-line Deduplication System

(Avani Wildani, Ethan L. Miller, Ohad Rodeh)

Holistic Data Cleaning: Putting Violations into Context

(Xu Chu, Ihab F. Ilyas, Paolo Papotti)

Inferring Data Currency and Consistency for Conflict Resolution

(Wenfei Fan, Floris Geerts, Nan Tang, Wenyuan Yu)

Research Session 14: Social Media I

LSII: An Indexing Structure for Exact Real-Time Search on Microblogs

(Lingkun Wu, Wenqing Lin, Xiaokui Xiao, Yabo Xu)

Utilizing Users' Tipping Points in E-Commerce Recommender Systems

(Kailun Hu, Wynne Hsu, Mong Li Lee)

Presenting Diverse Location Views with Real-Time Near-Duplicate Photo Elimination

(Jiajun Liu, Zi Huang, Hong Cheng, Yueguo Chen, Heng Tao Shen, Yanchun Zhang)

Research Session 15: Data Trust

Publicly Verifiable Grouped Aggregation Queries on Outsourced Data Streams

(Suman Nath, Ramarathnam Venkatesan)

Trustworthy Data from Untrusted Databases

(Rohit Jain, Sunil Prabhakar)

On the Relative Trust Between Inconsistent Data and Inaccurate Constraints

(George Beskales, Ihab F. Ilyas, Lukasz Golab, Artur Galiullin)

Research Session 16: Data on the Cloud

Catch the Wind: Graph Workload Balancing on Cloud

(Zechao Shang, Jeffrey Xu Yu)

EAGRE: Towards Scalable I/O Efficient SPARQL Query Evaluation on the Cloud

(Xiaofei Zhang, Lei Chen, Yongxin Tong, Min Wang)

C-Cube: Elastic Continuous Clustering in the Cloud

(Zhenjie Zhang, Hu Shu, Zhihong Chong, Hua Lu, Yin Yang)

Research Session 17: Similarity Ranking

Efficient Search Algorithm for SimRank

(Yasuhiro Fujiwara, Makoto Nakatsuji, Hiroaki Shiokawa, Makoto Onizuka)

Towards Efficient SimRank Computation on Large Networks

(Weiren Yu, Xuemin Lin, Wenjie Zhang)

RoundTripRank: Graph-Based Proximity with Importance and Specificity

(Yuan Fang, Kevin Chen-Chuan Chang, Hady W. Lauw)

Research Session 18: Spatial Databases

Finding Distance-Preserving Subgraphs in Large Road Networks

(Da Yan, James Cheng, Wilfred Ng, Steven Liu)

Maximum Visibility Queries in Spatial Databases

(Sarah Masud, Farhana Murtaza Choudhury, Mohammed Eunus Ali, Sarana Nutanong)

Memory-Efficient Algorithms for Spatial Network Queries

(Sarana Nutanong, Hanan Samet)

Research Session 19: Social Media II

A Unified Model for Stable and Temporal Topic Detection from Social Media Data

(Hongzhi Yin, Bin Cui, Hua Lu, Yuxin Huang, Junjie Yao)

Crowdsourced Enumeration Queries

(Beth Trushkowsky, Tim Kraska, Michael J. Franklin, Purnamrita Sarkar)

On Incentive-Based Tagging

(Xuan S. Yang, Reynold Cheng, Luyi Mo, Ben Kao, David W. Cheung)

Research Session 20: Trees and XML

Ontology-Based Subgraph Querying

(Yinghui Wu, Shengqi Yang, Xifeng Yan)

Stratification Driven Placement of Complex Data: A Framework for Distributed Data Analytics

(Ye Wang, Srinivasan Parthasarathy, P. Sadayappan)

Optimizing Approximations of DNF Query Lineage in Probabilistic XML

(Asma Souihli, Pierre Senellart)

Research Session 21: Security and Privacy

Secure Nearest Neighbor Revisited

(Bin Yao, Feifei Li, Xiaokui Xiao)

Accurate and Efficient Private Release of Datacubes and Contingency Tables

(Grigory Yaroslavtsev, Graham Cormode, Cecilia M. Procopiuc, Divesh Srivastava)

Differentially Private Grids for Geospatial Data

(Wahbeh Qardaji, Weining Yang, Ninghui Li)

Research Session 22: Randomized Algorithms for Graphs

Faster Random Walks by Rewiring Online Social Networks On-the-Fly

(Zhuojie Zhou, Nan Zhang, Zhiguo Gong, Gautam Das)

Sampling Node Pairs Over Large Graphs

(Pinghui Wang, Junzhou Zhao, John C.S. Lui, Don Towsley, Xiaohong Guan)

Link Prediction Across Networks by Biased Cross-Network Sampling

(Guo-Jun Qi, Charu C. Aggarwal, Thomas Huang)

Research Session 23: Distributed Data Processing

Interval Indexing and Querying on Key-Value Cloud Stores

(George Sfakianakis, Ioannis Patlakas, Nikos Ntarmos, Peter Triantafillou)

Robust Distributed Stream Processing

(Chuan Lei, Elke A. Rundensteiner, Joshua D. Guttman)

Research Session 24: Data Mining II

Learning to Rank from Distant Supervision: Exploiting Noisy Redundancy for Relational Entity Search

(Mianwei Zhou, Hongning Wang, Kevin Chen-Chuan Chang)

AFFINITY: Efficiently Querying Statistical Measures on Time-Series Data

(Saket Sathe, Karl Aberer)

Forecasting the Data Cube: A Model Configuration Advisor for Multi-Dimensional Data Sets

(Ulrike Fischer, Christopher Schildt, Claudio Hartmann, Wolfgang Lehner)

Research Session 25: Lineage and Provenance

SubZero: A Fine-Grained Lineage System for Scientific Databases

(Eugene Wu, Samuel Madden, Michael Stonebraker)

Logical Provenance in Data-Oriented Workflows

(Robert Ikeda, Akash Das Sarma, Jennifer Widom)

Revision Provenance in Text Documents of Asynchronous Collaboration

(Jing Zhang, H.V. Jagadish)

Research Session 26: Similarity Search

Inverted Linear Quadtree: Efficient Top K Spatial Keyword Search

(Chengyuan Zhang, Ying Zhang, Wenjie Zhang, Xuemin Lin)

Similarity Query Processing for Probabilistic Sets

(Ming Gao, Cheqing Jin, Wei Wang, Xuemin Lin, Aoying Zhou)

Top-k String Similarity Search with Edit-Distance Constraints

(Dong Deng, Guoliang Li, Jianhua Feng, Wen-Syan Li)

Research Session 27: Shortest and Direct Query

On Shortest Unique Substring Queries

(Jian Pei, Wush Chi-Hsuan Wu, Mi-Yen Yeh)

Engineering Generalized Shortest Path Queries

(Michael N. Rice, Vassilis J. Tsotras)

Efficient Direct Search on Compressed Genomic Data

(Xiaochun Yang, Bin Wang, Chen Li, Jiaying Wang, Xiaohui Xie)

Research Session 28: Skyline and Snapshot Query

On Answering Why-Not Questions in Reverse Skyline Queries

(Md. Saiful Islam, Rui Zhou, Chengfei Liu)

Layered Processing of Skyline-Window-Join (SWJ) Queries Using Iteration-Fabric

(Mithila Nagendra, K. Selçuk Candan)

Efficient Snapshot Retrieval Over Historical Graph Data

(Udayan Khurana, Amol Deshpande)

Research Session 29: Large Graph Indexing

FERRARI: Flexible and Efficient Reachability Range Assignment for Graph Indexing

(Stephan Seufert, Avishek Anand, Srikanta Bedathur, Gerhard Weikum)

gIceberg: Towards Iceberg Analysis in Large Graphs

(Nan Li, Ziyu Guan, Lijie Ren, Jian Wu, Jiawei Han, Xifeng Yan)

Top-k Graph Pattern Matching Over Large Graphs

(Jiefeng Cheng, Xianggang Zeng, Jeffrey Xu Yu)

Research Session 30: Web Data

Breaking the Top-k Barrier of Hidden Web Databases

(Saravanan Thirumuruganathan, Nan Zhang, Gautam Das)

Automatic Extraction of Top-k Lists from the Web

(Zhixian Zhang, Kenny Q. Zhu, Haixun Wang, Hongsong Li)

Finding Interesting Correlations with Conditional Heavy Hitters

(Katsiaryna Mirylenka, Themis Palpanas, Graham Cormode, Divesh Srivastava)

Research Session 31: Query Optimization

Predicting Query Execution Time: Are Optimizer Cost Models Really Unusable?

(Wentao Wu, Yun Chi, Shenghuo Zhu, Junichi Tatemura, Hakan Hacigümüs, Jeffrey F. Naughton)

Query Optimization for Differentially Private Data Management Systems

(Shangfu Peng, Yin Yang, Zhenjie Zhang, Marianne Winslett, Yong Yu)

Top Down Plan Generation: From Theory to Practice

(Pit Fender, Guido Moerkotte)

Research Session 32: Data Storage

TBF: A Memory-Efficient Replacement Policy for Flash-Based Caches

(Cristian Ungureanu, Biplob Debnath, Stephen Rago, Akshat Aranya)

Fast Peak-to-Peak Behavior with SSD Buffer Pool

(Jaeyoung Do, Donghui Zhang, Jignesh M. Patel, David J. DeWitt)

SELECT Triggers for Data Auditing

(Daniel Fabbri, Ravi Ramamurthy, Raghav Kaushik)

Industry Session 1

Data Services for E-Tailers: Leveraging Web Search Engine Assets

(Tao Cheng, Kaushik Chakrabarti, Surajit Chaudhuri, Vivek Narasayya, Manoj Syamala)

SAP HANA Distributed In-Memory Database System: Transaction, Session, and Metadata Management

(Juchang Lee, Yong Sik Kwon, Franz Färber, Michael Muehle, Chulwon Lee, Christian Bensberg, Joo Yeon Lee, Arthur H. Lee, Wolfgang Lehner)

Industry Session 2

HFMS: Managing the Lifecycle and Complexity of Hybrid Analytic Data Flows

(Alkis Simitsis, Kevin Wilkinson, Umeshwar Dayal, Meichun Hsu)

KuaFu: Closing the Parallelism Gap in Database Replication

(Chuntao Hong, Dong Zhou, Mao Yang, Carbo Kuo, Lintao Zhang, Lidong Zhou)

Materialization Strategies in the Vertica Analytic Database: Lessons Learned

(Lakshmikant Shrinivas, Sreenath Bodagala, Ramakrishna Varadarajan, Ariel Cary, Vivek Bharathan, Chuck Bear)

Industry Session 3

Pipe Failure Prediction: A Data Mining Method

(Rui Wang, Weishan Dong, Yu Wang, Ke Tang, Xin Yao)

SASH: Enabling Continuous Incremental Analytic Workflows on Hadoop

(Manish Sethi, Narendran Sachindran, Sriram Raghavan)

Automating Pattern Discovery for Rule Based Data Standardization Systems

(Snigdha Chaturvedi, Hima Prasad K., Tanveer A. Faruque, Bhupesh S. Chawda, L. Venkata Subramaniam, Raghu Krishnapuram)

Seminars

Machine Learning on Big Data

(Tyson Condie, Paul Mineiro, Neoklis Polyzotis, Markus Weimer)

Big Data Integration

(Xin Luna Dong, Divesh Srivastava)

Workload Management for Big Data Analytics

(Ashraf Aboulnaga, Shivnath Babu)

Knowledge Harvesting from Text and Web Sources

(Fabian Suchanek, Gerhard Weikum)

Sorting in Space: Multidimensional, Spatial, and Metric Data Structures for Applications in Spatial Databases, Geographic Information Systems (GIS), and Location-Based Services

(Hanan Samet)

Triples in the Clouds

(Zoi Kaoudi, Ioana Manolescu)

Querying Encrypted Data

(Arvind Arasu, Ken Eguro, Raghav Kaushik, Ravi Ramamurthy)

Shallow Information Extraction for the Knowledge Web

(Denilson Barbosa, Haixun Wang, Cong Yu)

Secure and Privacy-Preserving Database Services in the Cloud

(Divyakant Agrawal, Amr El Abbadi, Shiyuan Wang)

Demo Groups 1 and 2

Twitter+: Build Personalized Newspaper for Twitter

(Chen Liu, Anthony K.H. Tung)

A Generic Database Benchmarking Service

(Martin Kaufmann, Peter M. Fischer, Donald Kossmann, Norman May)

Aeolus: An Optimizer for Distributed Intra-Node-Parallel Streaming Systems

(Matthias J. Sax, Malu Castellanos, Qiming Chen, Meichun Hsu)

Crowd-Answering System via Microblogging

(Xianke Zhou, Ke Chen, Sai Wu, Bingbing Zhang)

With a Little Help from My Friends

(Arnab Nandi, Stelios Pappas, John C. Shafer, Rakesh Agrawal)

Peeking into the Optimization of Data Flow Programs with MapReduce-Style UDFs

(Fabian Hueske, Mathias Peters, Aljoscha Krettek, Matthias Ringwald, Kostas Tzoumas, Volker Markl, Johann-Christoph Freytag)

Very Fast Estimation for Result and Accuracy of Big Data Analytics: The EARL System

(Nikolay Laptev, Kai Zeng, Carlo Zaniolo)

Road Network Mix-Zones for Anonymous Location Based Services

(Balaji Palanisamy, Sindhuja Ravichandran, Ling Liu, Binh Han, Kisung Lee, Calton Pu)

Query Time Scaling of Attribute Values in Interval Timestamped Databases

(Anton Dignös, Michael Böhlen, Johann Gamper)

Extracting Interesting Related Context-Dependent Concepts from Social Media Streams Using Temporal Distributions

(Craig P. Sayers, Meichun Hsu)

Demo Groups 1 and 2 continues next page ...

Demo Groups 1 and 2 continued ...

VERDICT: Privacy-Preserving Authentication of Range Queries in Location-Based Services

(Haibo Hu, Qian Chen, Jianliang Xu)

ExpFinder: Finding Experts by Graph Pattern Matching

(Wenfei Fan, Xin Wang, Yinghui Wu)

Tajo: A Distributed Data Warehouse System on Large Clusters

(Hyunsik Choi, Jihoon Son, Haemi Yang, Hyoseok Ryu, Byungnam Lim, Soohyung Kim, Yon Dohn Chung)

Demo Groups 3 and 4

Igora: An Integration System for Probabilistic Data

(Dan Olteanu, Lampros Papageorgiou, Sebastiaan J. van Schaik)

Complex Pattern Matching in Complex Structures: The XSeq Approach

(Kai Zeng, Mohan Yang, Barzan Mozafari, Carlo Zaniolo)

T-Music: A Melody Composer Based on Frequent Pattern Mining

(Cheng Long, Raymond Chi-Wing Wong, Raymond Ka Wai Sze)

SHARE: Secure Information Sharing Framework for Emergency Management

(Barbara Carminati, Elena Ferrari, Michele Guglielmi)

KORS: Keyword-Aware Optimal Route Search System

(Xin Cao, Lisi Chen, Gao Cong, Jihong Guan, Nhan-Tue Phan, Xiaokui Xiao)

CrowdPlanr: Planning Made Easy with Crowd

(Iliia Lotosh, Tova Milo, Slava Novgorodov)

ASVTDECTOR: A Practical Near Duplicate Video Retrieval System

(Xiangmin Zhou, Lei Chen)

Demo Groups 3 and 4 continues next page ...

Demo Groups 3 and 4 continued ...

YumiInt — A Deep Web Integration System for Local Search Engines for Geo-Referenced Objects

(Eduard Dragut, B.P. Beirne, A. Neyestani, B. Atassi, Clement Yu, Bhaskar DasGupta, Weiyi Meng)

A Demonstration of the G* Graph Database System

(Sean R. Spillane, Jeremy Birnbaum, Daniel Bokser, Daniel Kemp, Alan Labouseur, Paul W. Olsen Jr., Jayadevan Vijayan, Jeong-Hyon Hwang, Jun-Weon Yoon)

RECODS: Replica Consistency-On-Demand Store

(Yuqing Zhu, Philip S. Yu, Jianmin Wang)

SODIT: An Innovative System for Outlier Detection Using Multiple Localized Thresholding and Interactive Feedback

(Ji Zhang, Hua Wang, Xiaohui Tao, Lili Sun)

COLA: A Cloud-Based System for Online Aggregation

(Yantao Gan, Xiaofeng Meng, Yingjie Shi)

RoadAlarm: A Spatial Alarm System on Road Networks

(Kisung Lee, Emre Yigitoglu, Ling Liu, Binh Han, Balaji Palanisamy, Calton Pu)

A Real-Time Abnormality Detection System for Intensive Care Management

(Guangyan Huang, Jing He, Jie Cao, Zhi Qiao, Michael Steyn, Kersi Taraporewalla)