

---

# Advances in Database Technology — EDBT 2020

23rd International Conference  
on Extending Database Technology  
Copenhagen, Denmark, March 30–April 2, 2020  
Proceedings

*Editors*

Angela Bonifati  
Yongluan Zhou  
Marcos Antonio Vaz Salles  
Alexander Böhm  
Dan Olteanu  
George Fletcher  
Arijit Khan  
Bin Yang



Advances in Database Technology – EDBT 2020  
Proceedings of the 23rd International Conference  
on Extending Database Technology  
Copenhagen, Denmark, March 30–April 2, 2020

Series ISSN: 2367-2005

*Editors*

Angela Bonifati, Lyon 1 University, France  
Yongluan Zhou, University of Copenhagen, Denmark  
Marcos Antonio Vaz Salles, University of Copenhagen, Denmark  
Alexander Böhm, SAP Research, Germany  
Dan Olteanu, University of Oxford, United Kingdom  
George Fletcher, Eindhoven University of Technology, The Netherlands  
Arijit Khan, Nanyang Technological University, Singapore  
Bin Yang, Aalborg University, Denmark



OpenProceedings.org  
University of Konstanz  
University Library  
78457 Konstanz, Germany

COPYRIGHT NOTICE: Copyright © 2020 by the authors of the individual papers.

Distribution of all material contained in this volume is permitted under the terms of the Creative Commons license CC-by-nc-nd 4.0

OpenProceedings ISBN: 978-3-89318-083-7

DOI of this front matter: 10.5441/002/edbt.2020.01

# Foreword

The International Conference on Extending Database Technology (EDBT) is a leading international forum for database researchers, developers, and users to present and discuss novel and cutting-edge ideas and techniques, and to showcase tools and experiences related to data management. Data management is an essential enabling technology that has applications in several scientific, business and social communities, and runs on diverse technical platforms associated with the web, enterprises, clouds and mobile devices. The database community has a continuing tradition of contributing with models, algorithms and architectures to the set of tools and applications that enable day-to-day functioning of our societies. Faced with the broad challenges of today's applications, data management technology constantly broadens its reach, exploiting new hardware and software to achieve innovative results and embracing new challenges in the years to come.

EDBT 2020 solicited submissions of original research contributions, descriptions of industrial solutions and applications, and proposals for tutorials and software demonstrations. We encouraged submissions of research papers related to all aspects of data management. We also encouraged submissions of visionary papers as well as innovative system papers and experimental analyses papers. In addition to long research paper submissions, EDBT 2020 again encouraged the submission of short research papers, which provide an excellent opportunity to describe significant work or research in progress that can foster the discussions at the conference. Short papers are presented as posters at plenary poster sessions of the conference. This year, they will also be communicated at a plenary lightning talks session.

The program committees of EDBT accepted 30 out of 151 submitted regular research papers, resulting in an acceptance rate of 20% for the research track; 26 out of 85 submitted short research papers, resulting in an acceptance rate of 31% for short research papers; 16 out of 37 demos, resulting in an acceptance rate of 43% for the demonstration track; and 10 out of 37 industrial and application papers, resulting in an acceptance rate of 27% for industrial and application papers. The papers will be presented in eight research paper sessions, three industrial and application sessions, as well as two plenary poster and demonstration sessions.

The program additionally features five workshops, one of which is the well-established DOLAP workshop that has successfully been co-located with EDBT since many years. Finally, the conference program includes four tutorials and an EDBT and ICDT joint session on climate change.

I would like to thank all authors for their contributions: a successful conference crucially depends on high-quality submissions. I also would like to thank all senior reviewers and reviewers for serving on the EDBT 2020 program committee, in particular for the high quality and timely handling of all reviews and discussions. This community service requires a lot of work on a tight schedule, and is what makes our research community function and ensures the sustained impact of our research. Thanks to their valuable effort we can look forward to an exciting program and attractive EDBT conference in Copenhagen from March 30–April 2, 2020.

A warm thanks to Anastasia Alaimaki and Tamer Özsu for serving on the Test-of-Time Award committee to select the paper from EDBT 2010 that has had the most lasting influence. Wook-shin Han, Erhard Rahm and Nesime Tatbul generously accepted to serve on the Best Paper committee. The EDBT 2020 program is the result of the joint effort of many people who shared their experience and time to contribute to the EDBT 2020 program and make the conference a great success. Alexander Boehm served as PC chair for industrial and application papers; George Fletcher as PC chair for the demonstration track; Dan Olteanu as tutorial chair; Alexandra Poulouvasilis as workshop chair. My warmest thanks to all these people. The general chairs, Yongluan Zhou and Marcos Vaz Salles and the other local organizers worked hard to make all necessary arrangements for a successful event. Special thanks to Arijt Khan, the EDBT proceedings chair; Davide Mottin, the publicity chair; Bin Yang, the local executive chair; Boris Düdder, the sponsorship chair; Nanna Højholt, the finance chair and Yiwen Wang, the website chair, for tirelessly finding solutions for all our requests. Norman Paton was most helpful in advising and coordinating with the EDBT Executive Board. Last but not least, I would like to thank Marc H. Scholl for assembling the EDBT proceedings on [openproceedings.org](http://openproceedings.org) I hope that you find EDBT 2020 inspiring, enriching, and enjoyable and look forward to meeting you in Copenhagen.

Angela Bonifati  
EDBT 2020 Program Chair

# Program Committee Members

## Research Program Committee

Angela Bonifati (Lyon 1 U, France) – Chair

### Senior Program Committee Members

Karl Aberer (EPFL Lausanne, Switzerland)	Ulf Leser (Humboldt-U Berlin, Germany)
Walid Aref (Purdue U, USA)	Guoliang Li (Tsinghua U, China)
Michael Benedikt (U Oxford, UK)	Chengkai Li (U Texas at Arlington, USA)
Michael Böhlen (U Zurich, Switzerland)	Eric Lo (Chinese U Hong Kong, China)
K. Selcuk Candan (Arizona State U, USA)	Evaggelia Pitoura (U Ioannina, Greece)
Kevin Chang (U Illinois at Urbana-Champaign, USA)	Louïqa Raschid (U Maryland, USA)
Vassilis Christophides (INRIA, France)	Sherif Sakr (U Tartu, Estonia)
Daniel Deutch (Tel Aviv U, Israel)	Semih Salihoglu (U Waterloo, Canada)
Floris Geerts (U Antwerp, Belgium)	Kai-Uwe Sattler (TU Ilmenau, Germany)
Jan Hidders (VU Brussel, Belgium)	Arash Termehchy (Oregon State U, USA)
Katja Hose (Aalborg U, Denmark)	Riccardo Torlone (Roma Tre U, Italy)
Christoph Koch (EPFL Lausanne, Switzerland)	Peter Triantafillou (U Warwick, UK)
Georgia Koutrika (Athena Research Center, Greece)	Yannis Velegarakis (Utrecht U, The Netherlands)

### Program Committee Members

Bernd Amann (Sorbonne U – LIP6, France)	Yaron Kanza (AT&T Labs – Research, USA)
Akhil Arora (EPFL Lausanne, Switzerland)	Asterios Katsifodimos (Delft UT, The Netherlands)
Elena Baralis (Politecnico di Torino, Italy)	Xiang Lian (Kent State U, USA)
Denilson Barbosa (U Alberta, Canada)	Ping Lu (Beihang U, China)
Senjuti Basu Roy (New Jersey Inst. of Techn., USA)	Paolo Missier (Newcastle U, USA)
Luigi Bellomarini (Banca d'Italia, Italy)	Davide Mottin (Aarhus U, Denmark)
Sonia Bergamaschi (U Modena Reggio Emilia, Italy)	Behrooz Omidvar-Tehrani (Grenoble Alpes, France)
Laure Berti-Equille (IRD, France)	Eric Peukert (Leipzig U, Germany)
Arnab Bhattacharya (IIT Kanpur, India)	Holger Pirk (Imperial College London, UK)
Luc Bouganim (INRIA-UVSQ, France)	Dimitris Plexousakis (Institute of CS, FORTH, Greece)
Andrea Cali (U London, Birkbeck College, UK)	Giuseppe Polese (U Salerno, Italy)
Bogdan Cautis (Paris Sud U, France)	Arnau Prat (U Politècn. Catalunya, Spain)
Lei Chen (Hong Kong U Sc. & Techn., China)	Mohammad Sadoghi (UC Davis)
Dario Colazzo (Paris Dauphine U, France)	Carlo Sartiani (U della Basilicata, Italy)
Bin Cui (Peking U, China)	Stefanie Scherzinger (OTH Regensburg, Germany)
Alfredo Cuzzocrea (U Calabria, Italy)	Petra Selmer (Neo4j, UK)
Sabrina De Capitani di Vimercati (U Milan, Italy)	Juan F. Sequeda (Capsenta Labs, USA)
Stefania Dumbrava (ENSIIE Rennes, France)	Hala Skaf-Molli (U Nantes, France)
Donatella Firmani (Roma Tre U, Italy)	Kostas Stefanidis (U Tampere, Finland)
Rainer Gemulla (U Mannheim, Germany)	Gábor Szárnyas (Budapest U Tech. & Eco., Hungary)
Paolo Guagliardo (U Edinburgh, UK)	Ernest Teniente (U Politècn. Catalunya, Spain)
Xi He (U Waterloo, Canada)	Jens Teubner (TU Dortmund, Germany)
Xin Huang (Hong Kong Baptist U, China)	Farouk Toumani (U Clermont Auvergne, France)
Zsolt Istvan (IMDEA Software Institute, Spain)	Anthony K. H. Tung (National U Singapore)
Panos Kalnis (King Abdullah UST, Saudi Arabia)	Wendy Hui Wang (Stevens Inst. of Techn., USA)
Vana Kalogeraki (Athens U Ec. & Busin., Greece)	Nikolay Yakovets (TU Eindhoven, The Netherlands)
Verena Kantere (U Ottawa, Canada)	Demetrios Zeinalipour (U Cyprus, Greece)

## **Industrial/Applications Program Committee**

Roy (Sudipto) Chowdhuri (Salesforce)  
Colin Florendo (Google)  
Prasanta Ghosh (Microsoft)  
Fisnik Kastrati (Huawei)  
Martin Kaufmann (Teradata)  
Justin Levandoski (Amazon)  
Stefan Mandl (Exasol)  
Norman May (SAP SE)  
Ismail Oukid (TU Dresden)  
Fatma Ozcan (IBM)  
Orestis Polychroniou (Amazon)  
Danica Porobic (Oracle)  
Harald Schöning (Software AG)  
Peter Thawley (Amazon)  
Adrian Vogelsgesang (Tableau)  
Hannes Voigt (Neo4j)  
Till Westmann (Couchbase)  
Panfeng Zhou (Alibaba)

## **Demonstration Program Committee**

Ashvin Agrawal (Microsoft Research, USA)  
Toshiyuki Amagasa (U Tsukuba, Japan)  
Sihem Amer-Yahia (CNRS LIG, France)  
Sourav Bhowmick (Nanyang Techn. U, Singapore)  
Carsten Binnig (TU Darmstadt, Germany)  
Chee-Yong Chan (National U Singapore)  
Adriane Chapman (U Southampton, UK)  
Dong-Wan Choi (Inha U, South Korea)  
Rick Cole (Tableau Software, Inc., USA)  
Vasiliki Kalavri (ETH Zurich, Switzerland)  
Romans Kasperovics (SAP, Germany)  
Wolfgang Lehner (TU Dresden, Germany)  
Hannes Mühleisen (CWI, Netherlands)  
Hubert Naacke (Sorbonne U, France)  
Keisuke Nakano (U Tohoku, Japan)  
Jorge Pérez (U Chile)  
Antonella Poggi (Sapienza U Roma, Italy)  
Cristian Riveros (Pontificia U Catolica Chile)  
Jagan Sankaranarayanan (Google, USA)  
Yuya Sasaki (Osaka U, Japan)  
Michael Schmidt (Amazon, USA)  
Sławek Staworko (U Lille 3, France)  
Dimitri Surinx (Hasselt U, Belgium)  
Alex Thomo (U Victoria, Canada)  
Xiaolan Wang (Megagon Labs, USA)  
Raymond Chi-Wing Wong (Hong Kong UST)  
Meihui Zhang (Beijing Inst. Techn., China)

# Conference Organization

## **General Chairs**

Yongluan Zhou, University of Copenhagen, Denmark  
Marcos Antonio Vaz Salles, University of Copenhagen, Denmark

## **EDBT Program Chair**

Angela Bonifati, Lyon 1 University, France

## **ICDT Program Chair**

Carsten Lutz, University of Bremen, Germany

## **EDBT Industrial/Application Chair**

Alexander Böhm, SAP Research, Germany

## **EDBT Demonstrations Chair**

George Fletcher, Eindhoven University of Technology, The Netherlands

## **Tutorial Chair**

Dan Olteanu, University of Oxford, United Kingdom

## **Workshops Chair**

Alex Poulouvasilis, Birbeck University of London, United Kingdom

## **EDBT Proceedings Chair**

Arijt Khan, Nanyang Technological University, Singapore

## **ICDT Proceedings Chair**

Jean Christoph Jung, University of Bremen, Germany

## **Local Executive Chair**

Bin Yang, Aalborg University, Denmark

## **Sponsorship Chairs**

Boris Düdler, University of Copenhagen, Denmark

## **Publicity Chair**

Davide Mottin, Aarhus University, Denmark

## **Finance Chair**

Nanna Højholt, University of Copenhagen, Denmark

## **Website Chair**

Yiwen Wang, University of Copenhagen, Denmark

# Test-of-Time Award

Established in 2014, the Test-of-Time Award awarded by the Extended Database Technology (EDBT) Conference recognizes papers presented at EDBT Conferences that have had the most impact in terms of research, methodology, conceptual contribution, or transfer to practice over the past ten years.

The EDBT 2020 Test of Time Award committee was formed by Anastasia Ailamaki (Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland), Tamer Oszu (University of Waterloo, Canada), and Angela Bonifati (Lyon 1 University, France).

After careful consideration, the committee has decided to select the following paper from the EDBT 2010 conference as the EDBT ToT Award winner for 2020:

## **Optimizing joins in a map-reduce environment**

by Foto Afrati and Jeff Ullman

published in EDBT 2010 Proceedings, pp. 99–110, DOI: 10.1145/1739041.1739056.

This paper presented optimization strategies for executing multi-way joins in a map-reduce environment. It focused on large-scale data and provided algorithms to choose the number of map-keys and shares in order to minimize the communication cost among the map and reduce processes.

The committee members agreed that this paper clearly pioneered the field of join processing in map-reduce environments. It has triggered substantial follow-up research and impact on big data processing in parallel and distributed architectures.

The EDBT Test-of-Time award for 2020 will be presented during the EDBT/ICDT 2020 Conference in Copenhagen, Denmark, as part of the Awards session on March 31, 2020.

# Best Paper Award

The EDBT 2020 Best Paper Award committee was formed by Wook-shin Han (Postech, Korea), Erhard Rahm (University of Leipzig, Germany), Nesime Tatbul (Intel & MIT, USA), and Angela Bonifati (Lyon 1 University, France). After careful consideration, the committee has decided to select the following paper as the EDBT Best Paper for 2020:

## **Provenance for Probabilistic Logic Programs**

by Shaobo Wang, Hui Lyu, Jaichi Zhang, Chenyuan Wu, Xinyi, Chen, Wenchao, Zhou, Boon Thau Loo, Susan B. Davidson, Chen Chen.

DOI: 10.5441/002/edbt.2020.14

**Abstract:** Despite the emergence of probabilistic logic programming (PLP) languages for data driven applications, there are currently no debugging tools based on provenance for PLP programs. In this paper, we propose a novel provenance model and system, called P3 (Provenance for Probabilistic logic Programs) for analyzing PLP programs. P3 enables four types of provenance queries: traditional explanation queries, queries for finding the set of most important derivations within an approximate error, top-K most influential queries, and modification queries that enable us to modify tuple probabilities with fewest modifications to program or input data. We apply these queries into real-world scenarios and present theoretical analysis and practical algorithms for such queries. We have developed a prototype of P3, and our evaluation on real-world data demonstrates that the system can support a wide-range of provenance queries with explainable results. Moreover, the system maintains provenance and execute queries efficiently with low overhead.

The EDBT Best Paper Awards for 2020 will be presented during the EDBT/ICDT 2020 Conference in Copenhagen, Denmark, as part of the Awards session on March 31, 2020.



# Table of Contents

Foreword	i
Program Committee Members	ii
Conference Organization	iv
Test-of-Time Award	v
Best Paper Award	vi
Table of Contents	vii
<b>Research Papers</b>	
Automatic Canonical Utterance Generation for Task-Oriented Bots from API Specifications <i>Mohammad-Ali Yaghoub-Zadeh-Fard, Boualem Benatallah, Shayan Zamanirad</i>	1
Efficient Continuous Multi-Query Processing over Graph Streams <i>Lefteris Zervakis, Vinay Setty, Christos Tryfonopoulos, Katja Hose</i>	13
Zooming Out on an Evolving Graph <i>Amir Aghasadeghi, Vera Moffitt, Sebastian Schelter, Julia Stoyanovich</i>	25
Efficient Search for Multi-Scale Time Delay Correlations in Big Time Series Data <i>Nguyen Ho, Torben Bach Pedersen, Van Long Ho, Mai Vu</i>	37
Dynamic Query Refinement for Interactive Data Exploration <i>Alexander Kalinin, Ugur Cetintemel, Zheguang Zhao, Stanley Zdonik</i>	49
Micro Analysis to Enable Energy-Efficient Database Systems <i>Chen Yang, Yongjie Du, Zhihui Du, Meng Xiaofeng</i>	61
Q-Store: Distributed, Multi-partition Transactions via Queue-oriented Execution and Communication <i>Thamir Qadah, Suyash Gupta, Mohammad Sadoghi</i>	73
Ensemble Grammar Induction For Detecting Anomalies in Time Series <i>Yifeng Gao, Jessica Lin, Constantin Brif</i>	85
Incremental Based Framework for Efficient Top-K Similarity Search in Interactive Data Analysis Sessions <i>Amit Somech, Tova Milo, Oded Elbaz</i>	97
pgFMU: Integrating Data Management with Physical System Modelling <i>Olga Rybnytska, Laurynas Šikšnys, Torben Bach Pedersen, Bijay Neupane</i>	109
Differentially-Private Next-Location Prediction with Neural Networks <i>Ritesh Ahuja, Gabriel Ghinita, Cyrus Shahabi</i>	121
Explaining Differences Between Unaligned Table Snapshots <i>Manuel Fink, Christian Meilicke, Heiner Stuckenschmidt</i>	133
Provenance for Probabilistic Logic Programs <i>Shaobo Wang, Hui Lyu, Jiachi Zhang, Chenyuan Wu, Xinyi Chen, Wenchao Zhou, Boon Thau Loo, Susan B. Davidson, Chen Chen</i>	145
Improved Cardinality Estimation by Learning Queries Containment Rates <i>Rojeh Hayek, Oded Shmueli</i>	157
Automated Management of Indexes for Dataflow Processing Engines in IaaS Clouds <i>Herald Kllapi, Ilija Pietri, Verena Kantere, Yannis Ioannidis</i>	169

Optimal Histograms with Outliers <i>Rachel Behar, Sara Cohen</i> . . . . .	181
BalanSiNG: Fast and Scalable Generation of Realistic Signed Networks <i>Jinhong Jung, Ha-Myung Park, U Kang</i> . . . . .	193
Distributed Similarity Joins over Top-K Rankings <i>Evica Milchevski, Sebastian Michel</i> . . . . .	205
MV-PBT: Multi-Version Indexing for Large Datasets and HTAP Workloads <i>Christian Riegger, Tobias Vincon, Robert Gottstein, Ilia Petrov</i> . . . . .	217
Lineage-Preserving Anonymization of the Provenance of Collection-Based Workflows <i>Khalid Belhajjame</i> . . . . .	229
Sharing Computations for User-Defined Aggregate Functions <i>Chao Zhang, Farouk Toumani</i> . . . . .	241
Tracing nested data with structural provenance for big data analytics <i>Ralf Diestelkämper, Melanie Herschel</i> . . . . .	253
A Parallel and Distributed Approach for Diversified Top-k Best Region Search <i>Hamid Shahrivari, Matthaios Olma, Odysseas Papapetrou, Dimitrios Skoutas, Anastasia Ailamaki</i> . . . . .	265
Data Curation with Deep Learning <i>Saravanan Thirumuruganathan, Nan Tang, Mourad Ouzzani, Anhai Doan</i> . . . . .	277
Fairness in Clustering with Multiple Sensitive Attributes <i>Savitha Abraham, Deepak P, Sowmya S Sundaram</i> . . . . .	287
Ontology-Based RDF Integration of Heterogeneous Data <i>Maxime Buron, Francois Goasdoue, Ioana Manolescu, Marie-Laure Mugnier</i> . . . . .	299
Manually Detecting Errors for Data Cleaning Using Adaptive Crowdsourcing Strategies <i>Haojun Zhang, Chengliang Chai, Anhai Doan, Paris Koutris, Esteban Arcaute</i> . . . . .	311
Publishing Video Data with Indistinguishable Objects <i>Han Wang, Yuan Hong, Yu Kong, Jaideep Vaidya</i> . . . . .	323
Efficient PrefDiv Algorithms for Effective Top-k Result Diversification <i>Xiaoyu Ge, Panos Chrysanthis</i> . . . . .	335
Adaptive Main-Memory Indexing for High-Performance Point-Polygon Joins <i>Andreas Kipf, Harald Lang, Varun Pandey, Raul Alexandru Persa, Christoph Anneser, Eleni Tzirita Zacharatou, Harish Doraiswamy, Peter Boncz, Thomas Neumann, Alfons Kemper</i> . . . . .	347
<b>Short papers</b>	
DISGD: A Distributed Shared-nothing Matrix Factorization for Large Scale Online Recommender Systems <i>Heidy Hazem, Ahmed Awad, Ahmed Hassan, Sherif Sakr</i> . . . . .	359
Forming Compatible Teams in Signed Networks <i>Ioannis Kouvatis, Konstantinos Semertzidis, Maria Zerva, Evaggelia Pitoura, Panayiotis Tsaparas</i> . . . . .	363
A Learning Based Approach to Predict Shortest-Path Distances <i>Jianzhong Qi, Wei Wang, Rui Zhang, Zhuowei Zhao</i> . . . . .	367
Optimizing Data Movement with Near-Memory Acceleration of In-memory DBMS <i>Donghun Lee, Andrew Chang, Minseon Ahn, Jongmin Gim, Jungmin Kim, Jaemin Jung, Kangwoo Choi, Vincent Pham, Oliver Reibholz, Krishna Malladi, Yang-Seok Ki</i> . . . . .	371
ODSA: Open Database Storage Access <i>James Wagner, Alexander Rasin, Dai Hai Ton That, Tanu Malik, Jonathan Grier</i> . . . . .	375

An Integrated Graph Neural Network for Supervised Non-obvious Relationship Detection in Knowledge Graphs <i>Phillipp Müller, Xiao Qin, Balaji Ganesan, Nasrullah Sheikh, Berthold Reinwald</i> . . . . .	379
Outlier detection in multivariate functional data based on a geometric aggregation <i>Clément Lejeune, Josiane Mothe, Olivier Teste</i> . . . . .	383
REMI: Mining Intuitive Referring Expressions on Knowledge Bases <i>Luis Galárraga, Julien Delaunay, Jean-Louis Dessalles</i> . . . . .	387
The Case for Hybrid Succinct Data Structures <i>Christoph Anneser, Andreas Kipf, Harald Lang, Thomas Neumann, Alfons Kemper</i> . . . . .	391
FairPrep: Promoting Data to a First-Class Citizen in Studies on Fairness-Enhancing Interventions <i>Sebastian Schelter, Yuxuan He, Jatin Khilnani, Julia Stoyanovich</i> . . . . .	395
Partially Materializable Delta Trees for Efficient Data Wrangling of Semi-Structured Contents <i>Nico Schäfer, Sebastian Michel</i> . . . . .	399
Towards Fine-Grained Data Access Control Through Active Peer Probing <i>Yael Amsterdamer, Osnat Drien</i> . . . . .	403
The ML-Index: A Multidimensional, Learned Index for Point, Range, and Nearest-Neighbor Queries <i>Angjela Davitkova, Evica Milchevski, Sebastian Michel</i> . . . . .	407
Retro: Relation Retrofitting For In-Database Machine Learning on Textual Data <i>Michael Günther, Maik Thiele, Wolfgang Lehner</i> . . . . .	411
Revisiting the Theory and Practice of Database Cracking <i>Fatemeh Zardbani, Peyman Afshani, Panagiotis Karras</i> . . . . .	415
Boosting Blocking Performance in Entity Resolution Pipelines: Comparison Cleaning using Bloom Filters <i>Leonardo Gazzarri, Melanie Herschel</i> . . . . .	419
Disco: Efficient Distributed Window Aggregation <i>Lawrence Benson, Philipp M. Grulich, Steffen Zeuch, Volker Markl, Tilmann Rabl</i> . . . . .	423
Explaining Missing Query Results in Natural Language <i>Daniel Deutch, Nave Frost, Amir Gilad, Tomer Haimovich</i> . . . . .	427
A Context-based Approach for Partitioning Big Data <i>Sara Migliorini, Alberto Belussi, Elisa Quintarelli, Damiano Carra</i> . . . . .	431
SlideSide: A fast Incremental Stream Processing Algorithm for Multiple Queries <i>Georgios Theodorakis, Peter Pietzuch, Holger Pirk</i> . . . . .	435
Efficient Enumeration of Four Node Graphlets at Trillion-Scale <i>Yudi Santoso, Venkatesh Srinivasan, Alex Thomo</i> . . . . .	439
Band Joins for Interval Data <i>Panagiotis Bouros, Konstantinos Lampropoulos, Dimitrios Tsitsigkos, Nikos Mamoulis, Manolis Terrovitis</i> . . . . .	443
Personalized Page Rank on Knowledge Graphs: Particle Filtering is all you need! <i>Denis Gallo, Matteo Lissandrini, Yannis Velegrakis</i> . . . . .	447
Towards Planning of Regular Queries with Memory <i>Thomas Mulder, Nikolay Yakovets, George Fletcher</i> . . . . .	451
Accurate Demand Forecasting for Retails with Deep Neural Networks <i>Shanhe Liao, Weixiong Rao</i> . . . . .	455
Efficient Skyline Computation in High-Dimensionality Domains <i>Rui Liu, Dominique Li</i> . . . . .	459

## Industry and Applications Papers

Entity Matching with Transformer Architectures - A Step Forward in Data Integration <i>Ursin Brunner, Kurt Stockinger</i> . . . . .	463
Gallery: A Machine Learning Model Management System at Uber <i>Chong Sun, Nader Azari, Chintan Turakhia</i> . . . . .	474
Diverse User Selection for Opinion Procurement <i>Yael Amsterdamer, Oded Goldreich</i> . . . . .	486
Elastic Scaling in VectorH <i>Steffen Kläbe, Kai-Uwe Sattler, Stephan Baumann, Michael Rink</i> . . . . .	498
Fairness in Online Jobs: A Case Study on TaskRabbit and Google <i>Sihem Amer-Yahia, Shady Elbassuoni, Ahmad Ghizzawi, Ria Mae Borromeo, Emilie Hoareau, Philippe Mulhem</i> . .	510
Inventory Reduction via Maximal Coverage in E-Commerce <i>Shay Gershtein, Tova Milo, Slava Novgorodov</i> . . . . .	522
Cost Estimation Across Heterogeneous SQL-Based Big Data Infrastructures in Teradata IntelliSphere <i>Kassem Awada, Mohamed Eltabakh, Conrad Tang, Mohammed Al-Kateb, Sanjay Nair, Grace Au</i> . . . . .	534
Fast Entropy Maximization for Selectivity Estimation of Conjunctive Predicates on CPUs and GPUs <i>Diego Havenstein, Peter Lysakovski, Norman May, Guido Moerkotte, Gabriele Steidl</i> . . . . .	546
Weaving Enterprise Knowledge Graphs: The Case of Company Ownership Graphs <i>Paolo Atzeni, Luigi Bellomarini, Michela Iezzi, Emanuel Sallinger, Adriano Vlad</i> . . . . .	555
Expanding Query Answers on Medical Knowledge Bases <i>Chuan Lei, Vasilis Efthymiou, Rebecca Geis, Fatma Ozcan</i> . . . . .	567
<b>Demonstrations</b>	
VAP: A Visual Analysis Tool for Energy Consumption Spatio-temporal Pattern Discovery <i>Xiufeng Liu, Zhibin Niu, Yanyan Yang, Junqi Wu, Dawei Cheng, Xin Wang</i> . . . . .	579
Chronos: The Swiss Army Knife for Database Evaluations <i>Marco Vogt, Alexander Stiemer, Sein Coray, Heiko Schuldt</i> . . . . .	583
skyex: an R Package for Entity Linkage <i>Suela Isaj, Torben Bach Pedersen</i> . . . . .	587
Data Quality Checking for Machine Learning with MeSQuaL <i>Ugo Comignani, Noël Novelli, Laure Berti-Equille</i> . . . . .	591
MALOS: A Movement-Aware Location Selection System <i>Di Yang, Hui Li, Meng Wang, Dan Li, Jiangtao Cui</i> . . . . .	595
RRAMEN: An Interactive Tool for Evaluating Choices and Changes in Transportation Networks <i>Camila Ferreira Costa, Theodoros Chondrogiannis, Mario A. Nascimento, Panagiotis Bouros</i> . . . . .	599
JedAI3 : beyond batch, blocking-based Entity Resolution <i>George Papadakis, Leonidas Tsekouras, Emmanouil Thanos, Nikiforos Pittaras, Giovanni Simonini, Dimitrios Skoutas, Paul Isaris, George Giannakopoulos, Themis Palpanas, Manolis Koubarakis</i> . . . . .	603
RetroLive: Analysis of Relational Retrofitted Word Embeddings <i>Michael Günther, Maik Thiele, Erik Nikulski, Wolfgang Lehner</i> . . . . .	607
RULER: Scaling Up Record-level Matching Rules <i>Luca Gagliardelli, Giovanni Simonini, Sonia Bergamaschi</i> . . . . .	611

Schema Mapping Generation in the Wild: A Demonstration with Open Government Data <i>Lacramioara Mazilu, Nikolaos Konstantinou, Norman Paton, Alvaro A. A. Fernandes</i> . . . . .	615
EPIQUE: Extracting Meaningful Science Evolution Patterns from Large Document Archives <i>Ke Li, Hubert Naacke, Bernd Amann</i> . . . . .	619
Task-Tuning in Privacy-Preserving Crowdsourcing Platforms <i>Joris Dugu��peroux, Antonin Voyez, Tristan Allard</i> . . . . .	623
Scaling a Public Transport Monitoring System to Internet of Things Infrastructures <i>Haralampos Gavriilidis, Adrian Michalke, Laura Mons, Steffen Zeuch, Volker Markl</i> . . . . .	627
Governor: Operator Placement for a Unified Fog-Cloud Environment <i>Ankit Chaudhary, Steffen Zeuch, Volker Markl</i> . . . . .	631
Human-in-the-Loop Schema Inference for Massive JSON Datasets <i>Mohamed-Amine Baazizi, Cl��ment Berti, Dario Colazzo, Giorgio Ghelli, Carlo Sartiani</i> . . . . .	635
SQL Query Processing Using an Integrated FPGA-based Near-Data Accelerator in ReProVide <i>Lekshmi B.G., Andreas Becher, Klaus Meyer-Wegener, Stefan Wildermann, J��rgen Teich</i> . . . . .	639
<b>Tutorials</b>	
Declarative Languages for Big Streaming Data <i>Riccardo Tommasini, Sherif Sakr, Emanuele Della Valle, Hojjat Jafarpour</i> . . . . .	643
Entity Resolution: Past, Present and Yet-to-Come <i>George Papadakis, Ekaterini Ioannou, Themis Palpanas</i> . . . . .	647
Fairness in Rankings and Recommenders <i>Evaggelia Pitoura, Georgia Koutrika, Kostas Stefanidis</i> . . . . .	651
NoSQL Schema Evolution and Data Migration: State-of-the-Art and Opportunities <i>Uta St��rl, Meike Klettke, Stefanie Scherzinger</i> . . . . .	655