Advances in Database Technology
— EDBT 2019

22nd International Conference
on Extending Database Technology
Lisbon, Portugal, March 26–29, 2019
Proceedings

Editors
Melanie Herschel
Helena Galhardas
Berthold Reinwald
Irini Fundulaki
Carsten Binnig
Zoi Kaoudi
In memoriam

Christine Collet
1956–2019
Foreword

The International Conference on Extending Database Technology (EDBT) is an established and renowned forum for the exchange of the latest research results and advances in data management. This year, the 22nd edition of EDBT takes place in Lisbon, Portugal, from March 26 to March 29, 2019. It is jointly organized with the International Conference on Database Theory (ICDT). In a world where increasingly many aspects of our lives and society are data-driven, data management technology continues to broaden its reach and extends its tradition of contributing models, algorithms, and architectures to novel applications adapted to new hardware and software.

As in previous years, EDBT 2019 solicited contributions both on novel research results and on experience and analysis results that focus on a comprehensive and detailed performance evaluation. For the first time, EDBT 2019 further solicited papers that describe innovative systems as part of its main research track. We also continued the recently established short paper track, offering a forum to present research in progress and visionary ideas during plenary poster sessions of the conference. To complement the scientific program, EDBT further solicited demonstrations of research prototypes, descriptions of industrial and application achievements, and proposals for tutorials.

The EDBT 2019 program committee reviewed 157 full research papers, of which 36 were accepted. For short papers, 28 papers out of 122 were selected. Among the 24 submissions to the industry and application track, 8 papers were accepted. The 21 demonstrations presented at the conference were selected among 42 demonstration proposals. Finally, we accepted 3 out of 10 tutorials. All these contributions will be presented at the conference. The program additionally features five workshops, an EDBT/ICDT joint session on research challenges, and four invited EDBT/ICDT joint keynotes.

Shaping the exciting program of EDBT 2019 is the result of a large community effort, and I take this opportunity to thank all persons involved. First, I would like to thank all authors for their high-quality submissions and contributions. I also would like to thank all reviewers who served on the EDBT 2019 program committee and the chairs responsible for our different tracks, namely Berthold Reinwald (IBM, United States) who chaired the industrial and application track, Carsten Binnig (TU Darmstadt, Germany) who served as demonstration chair, our tutorial chair Irini Fundulaki (ICS FORTH, Greece), and Paolo Papotti (EUROCOM, France) who served as workshop chair. The special session on joint EDBT/ICDT research challenges was organized by Julia Stoyanovich (NYU, USA). I also thank Laura Haas (UMass Amherst, USA) and Alon Halevy, who generously accepted to serve on the Test of Time Award Committee. Many thanks also to Paolo Atzeni (Universita’ Roma Tre, Italy), Wei Wang (UNSW Sydney, Australia), and Jeffrey Xu Yu (Chinese University of Hong Kong) for serving on the Best Paper Award committee.

The conference would not have been possible without the tireless effort of the general chair Helena Galhardas (INESC-ID and IST, Universidade de Lisboa, Portugal) and the local organization team. Special thanks to the finance chair Manuel J. Fonseca (Universidade de Lisboa, Portugal), the local executive chairs José Borbinha and Luís Rodrigues, the sponsorship chairs João Garcia and Miguel Pardal, the publicity chair Paolo Romano, the student volunteers chair Hugo Nicolau (all from INESC-ID and IST, Universidade de Lisboa, Portugal), and the website chair António Higgs (INESC-ID, Portugal). These proceedings have been produced thanks to our proceedings chair Zoi Kaoudi (QCRI, Qatar). Norman Paton was most helpful in advising and coordinating with the EDBT Executive Board.

I really look forward to an interesting program and exciting conference on March 26–29, 2019 and to meeting you in Lisbon.

Melanie Herschel
EDBT 2019 Program Chair
Program Committee Members

Research Program Committee

Adriane Chapman, U Southampton
Ahmed Elmagarmid, QCRi HBKU
Alekh Jindal, Microsoft
Alessandro Campi, Politecnico di Milano
Alexander Böhm, SAP SE
Alfons Kemper, TU Munich
Alfredo Cuzzocrea, ICAR-CNR and U Calabria
Aamelie Marian, Rutgers U
Amr El Abbadi, UC Santa Barbara
Angela Bonifati, U Lyon 1
Anton Dignós, Free U Bozen-Bolzano
Antonios Deligiannis, TU Crete
Arash Termehchy, Oregon State U
Arnab Bhattacharya, IIT Kanpur
Avigdor Gal, Technion
Bernhard Seeger, U Marburg
Boris Glavic, Illinois IT
Chuan Lei, IBM Research-Almaden
Daisy Zhe Wang, U Florida
Daniele Dell’Aglio, U Lyon 1
Dario Colazzo, LAMSAD - Universit
Davide Martinenghi, Politecnico di Milano
Demetrios Zeinalipour-Yazti, U Cyprus
Elena Ferrari, U Insubria
Elisa Quintarelli, Politecnico di Milano
Eric Lo, Chinese U Hong Kong
Ernest Teniente, U Politècnica de Catalunya
Esther Pacitti, LIRMM and INRIA U Montpellier 2
Farouk Toumani Limos, B. Pascal U Clermont-Ferrand
Felix Naumann, HPI
Francois Goasdoue, U Rennes 1
George H. L. Fletcher, Eindhoven UT
Georgia Koutrika, Athena Research Center
Goetz Graefe, Google
Gottfried Vossen, ERCIS Muenster
Guoliang Li, Tsinghua U
Haruo Yokota, Tokyo Institute of Technology
Heiko Schultd, U Basel
Isabel Cruz, U Illinois Chicago
Jean-Marc Petit, U Lyon & INSA Lyon
Jian Pei, Simon Fraser U
Jianwen Su, UC Santa Barbara
Johann Gamper, Free U Bozen-Bolzano
Kai-Uwe Sattler, TU Ilmenau
Karl Aberer, EPF Lausanne
Katja Hose, Aalborg U
Klemens Böhmer, Karlsruhe I.T.
Kostas Magoutis, U Ioannina
Kostas Stefanidis, U Tampere
Marta Patino, U Politécnica de Madrid
Martin Theobald, U Luxembourg
Matthias Boehm, IBM
Matthias Renz, George Mason U
Maya Ramanath, IIT Delhi
Michael Grossniklaus, U Konstanz
Minos Garofalakis, TU Crete
Mohammad Sadoghi UC Davis
Nicoleta Preda, U Versailles
Nikos Mamoulis, U Ioannina
Nikos Pelekis, U Piraeus
Panagiotis Karras, Aarhus U
Paolo Missier, Newcastle U
Periklis Andritsos, U Toronto
Peter Triantafillou, U Warwick
Riccardo Torlone, U Roma Tre
Sabrina De Capitani di Vimercati, U Milan
Sherif Sakr, U New South Wales
Stefan DeBloch, TU Kaiserslautern
Tore Risch, Uppsala U
Vassilis Christophides, INRIA
Verena Kantere, U Geneva
Will Wu, Google
Wolfgang Lehner, TU Dresden
Xiaochun Yang, Northeastern U
Xiaofang Zhou, U Queensland
Xin Huang, Hong Kong Baptist U
Yang Cao, U Edinburgh
Yannis Theodoridis, U Piraeus
Yannis Velegrakis, U Trento
Yaron Kanza, AT&T Labs - Research
Yi-Leh Wu, Nat. Taiwan U Sc. & Tech.
Zhifeng Bao, RMIT U
Ziawasch Abedjan, TU Berlin
Industrial/Applications Program Committee

Alkis Simitsis, HP Labs
Amol Ghoting, LinkedIn
Andrey Balmin, Workday
Anil Goel, SAP Canada
Benjamin Schlegel, Oracle Labs
Calisto Zuzarte, IBM Toronto
Cliff Leung, Huawei
David Simmen, Apple
Martin Grund, Amazon AWS
Michael Rys, Microsoft
Nesreen Ahmed, Intel
Niketan Pansare, IBM
Norbert Ritter, U Hamburg
Paul Brown, Teradata Corp
Songting Chen, Tigergraph
Yannis Sismanis, Google

Demonstration Program Committee

Alexander Böhm, SAP SE
Arvind Arasu, Microsoft
Dirk Habich, TU Dresden
Jana Gicheva, Imperial College London
Manos Athanassoulis, Boston U
Martin Theobald, U Luxembourg
Michael Gertz, Heidelberg U
Rick Cole, Tableau Software
Sanjay Krishnan, U Chicago
Stefan Manegold, CWI
Sven Helmer, Free U Bozen-Bolzano
Thibault Sellam, Columbia U
Tilmann Rabl, TU Berlin
Uwe Roehm, U Sydney
Viktor Leis, TU Munich
Ziawasch Abedjan, TU Berlin
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 2019 Test-of-Time Award committee was formed by Laura Haas (University of Massachusetts, USA), Alon Halevy, and Melanie Herschel, the EDBT 2019 PC chair. The committee was charged with selecting a paper from the EDBT 2009 Proceedings.

After careful consideration, the Test-of-Time Award committee decided for the following paper from the 2009 EDBT Conference held in Saint Petersburg, Russia to receive the award:

Shore-MT: a scalable storage manager for the multicore era
by Ryan Johnson, Ippokratis Pandis, Nikos Hardavellas, Anastasia Ailamaki, and Babak Falsafi

The committee members agreed that this paper clearly stands out in terms of methodology, impact, and influence. It has catalyzed and enabled substantial follow-up research and has demonstrated its high relevance to industry.

Abstract:
Historically, database engines focused on the ability to efficiently overlap many requests over a small number processor cores, with I/O latencies and scalability as the main design driver. However, the advent of increasingly multicore hardware circa 2000 brought new challenges because concurrent transactions begin to stress the limits of the storage manager’s thread scalability by accessing its internal structures simultaneously and in large numbers. This EDBT 2009 paper shows the results of experiments running benchmarks on four (then and still) popular open-source storage managers (Shore, BerkeleyDB, MySQL, and PostgreSQL) on a multicore machine. The results show that all systems suffer from scalability bottlenecks at the storage engine level. From that research emerged Shore-MT, an open-source multithreaded and highly scalable storage manager, built with Shore as a base. We learned that designers should favor scalability over single-thread performance, and we identified several other key principles for architecting scalable storage engines.

Ten years later, Shore-MT work has concluded, although the system still serves as a research platform in the space. Meanwhile, research on transaction processing scalability continues to mature, the move to main-memory transaction processing and their higher TPS increased the need for scalable storage managers, while the popular open-source systems, such as MySQL and PostgreSQL, significantly improved their scalability. In particular, a significant amount of research and industrial developments in the ten years since the Shore-MT paper focused on improving the scalability of individual components of a storage manager, such as latches, the logging subsystem and access methods. This research was partly carried out by our research group as follow-on work, but other research groups and database vendors have made important contributions as well. Another significant amount of effort has focused on scalable concurrency control protocols, again both within and outside our research group. The knowledge that we have gained from building Shore-MT has been invaluable in maintaining scalability in this new, multi-dimensional ecosystem.

The EDBT Test-of-Time award for 2019 will be presented during the EDBT/ICDT 2019 Conference as part of the Awards session on Wednesday, March 27, 2019, by Anastasia Ailamaki (EPFL, Switzerland).
# Table of Contents

Foreword ................................................. i  
Program Committee Members .......................... ii  
Test-of-Time Award ..................................... iv  
Table of Contents ........................................ v  

## Research Papers

**Umzi: Unified Multi-Zone Indexing for Large-Scale HTAP**  
Chen Luo, Pinar Tozun, Yuanyuan Tian, Ronald Barber, Vijayshankar Raman, Richard Sidle .................. 1  

**A Highly Scalable Labelling Approach for Exact Distance Queries in Complex Networks**  
Muhammad Farhan, Qing Wang, Yu Lin, Brendan Mckay ................................................................. 13  

**Stratified Random Sampling over Streaming and Stored Data**  
Trong Nguyen, Ming-Hung Shih, Divesh Srivastava, Srikanta Tirthapura, Bojian Xu ....................... 25  

**Boosting SimRank with Semantics**  
Brit Youngmann, Tova Milo, Amit Somech .................. 37  

**Leveraging Bitmap Indexing for Subgraph Searching**  
David Luaces, José R.R. Viqueira, Tomás F. Pena, José M. Cotos ...................................................... 49  

**Spec-QP: Speculative Query Planning for Joins over Knowledge Graphs**  
Madhalika Mohanty, Maya Ramanath, Mohamed Yahya, Gerhard Weikum ........................................ 61  

**Iterative Estimation of Mutual Information with Error Bounds**  
Michael Vollmer, Klemens Böhm .......................... 73  

**Functional Geometric Monitoring for Distributed Streams**  
Vasileios Samoladas, Minos Garofalakis .................. 85  

**Efficient Window Aggregation with General Stream Slicing.**  
Jonas Traub, Philipp M. Grillich, Alejandro Rodriguez Cuellar, Sebastian Breß, Asterios Katifodimos, Tilmann Rabl, Volker Markl ................................................................. 97  

**An Efficient Sliding Window Approach for Approximate Entity Extraction with Synonyms**  
Jin Wang, Chunbin Lin, Mingda Li, Carlo Zaniolo .................. 109  

**Attendance Maximization for Successful Social Event Planning**  
Nikos Bikakis, Vana Kalogeraki, Dimitrios Gunopulos .......................... 121  

**GroupTravel: Customizing Travel Packages for Groups**  
Siham Amer-Yahia, Shady Elbassuoni, Behrooz Omidvar-Tehrani, Ria Mae Borromeo, Mehrdad Farokhnejad .... 133  

**SEP2P: Secure and Efficient P2P Personal Data Processing**  
Julien Loudet, Iulian Sandu-Popa, Luc Bouganim .................. 145  

**Indexing Trajectories for Travel-Time Histogram Retrieval**  
Robert Waury, Christian S. Jensen, Satoshi Koida, Yoshiharu Ishikawa, Chuan Xiao ....................... 157  

**BB-Tree: A practical and efficient main-memory index structure for multidimensional workloads**  
Stefan Sprenger, Patrick Schäfer, Ulf Leser .................. 169  

**Semantic and Influence aware k-Representative Queries over Social Streams**  
Yanhao Wang, Yuchen Li, Kian-Lee Tan .................. 181
RDF graph summarization: principles, techniques and applications
Haridimos Kondylakis, Dimitris Kotzinos, Ioana Manolescu .................................................. 433

Schemas And Types For JSON Data
Mohamed-Amine Baazizi, Dario Colazzo, Giorgio Ghelli, Carlo Sartiani ................................. 437

Influence Maximization Revisited: The State of the Art and the Gaps that Remain
Akhil Arora, Sainyam Galhotra, Sayan Ranu .............................................................................. 440

Industry and Applications Papers

Finding Meaningful Contrast Patterns for Quantitative Data
Rohan Khade, Jessica Lin, Nital Patel .......................................................................................... 444

Predicting “What is Interesting” by Mining Interactive-Data-Analysis Session Logs
Amit Somech, Tova Milo, Chai Ozeri ....................................................................................... 456

Hidden Layer Models for Company Representations and Product Recommendations
Katsiaryna Mirylenka, Paolo Scotton, Christoph Miksovíc, Jeff Dillon ................................. 468

Big POI data integration with Linked Data technologies

Executing Entity Matching End to End: A Case Study
Pradap Konda, Sanjay Seshadri, Elan Segarra, Brent Hueth, Anhai Doan ................................ 489

The Copernicus App Lab project: Easy Access to Copernicus Data
Konstantina Bereta, Herve Caumont, Ulrike Daniels, Erwin Goor, Manolis Koubarakis, Despina-Athanasia Pantazi, George Stamoulis, Sam Ubels, Valentijn Venus, Firman Wahyudi .......................................................... 501

Modeling and Building IoT Data Platforms with Actor-Oriented Databases
Yiwen Wang, Julio Cesar Dos Reis, Kasper Myrtue Borggren, Marcos Antonio Vaz Salles, Claudia Bauzer Medeiros, Yongluan Zhou ........................................................................................................ 512

KSQL: Streaming SQL Engine for Apache Kafka
Hojjat Jafarpour, Rohan Desai .................................................................................................. 524

Demonstrations

The Power of SQL Lambda Functions
Maximilian Schüle, Dimitri Vorona, Linnea Passing, Harald Lang, Alfons Kemper, Stephan Günnemann, Thomas Neumann .................................................................................................................. 534

MINARET: A Recommendation Framework for Scientific Reviewers
Sherif Sakr, Mohamed Ragab, Mohamed Maher, Ahmed Awad .................................................. 538

SCube: A Tool for Segregation Discovery
Alessandro Baroni, Salvatore Ruggieri ..................................................................................... 542

SparkTune: tuning Spark SQL through query cost modeling
Enrico Gallinucci, Matteo Golfarelli ......................................................................................... 546

HOTMapper: Historical Open Data Table Mapper
Henrique Varella Ehrenfried, Rudolf Eckelberg, Hamer Iboshi, Eduardo Todt, Daniel Weingaertner, Marcos Didonet Del Fabro ......................................................................................................................... 550

SmartML: A Meta Learning-Based Framework for Automated Selection and Hyperparameter Tuning for Machine Learning Algorithms
Mohamed Maher, Sherif Sakr .................................................................................................. 554
devUDF: Increasing UDF development efficiency through IDE Integration. It works like a PyCharm!
Mark Raasveldt, Pedro Holanda, Stefan Manegold .......................................................... 558

ML2SQL - Compiling a Declarative Machine Learning Language to SQL and Python
Maximilian Schüle, Matthias Bünzroth, Dimitri Vorona, Alfons Kemper, Stephan Günnewann, Thomas Neumann

Incremental structural summarization of RDF graphs
François Goasdoue, Paweł Guzewicz, Ioana Manolescu .......................................................... 566

VISTA: A visual analytics platform for semantic annotation of trajectories
Amilcar Soares, Jordan Rose, Mohammad Elmemad, Chiara Renso, Stan Matwin .......................... 570

SLIPO: Large-Scale Data Integration for Points of Interest
Spiros Athanasiou, Michail Alexakis, Giorgos Giannopoulos, Nikos Karagiannakis, Yannis Kouvaras, Pantelis Mitropoulos, Kostas Patermpas, Dimitrios Skoutas .......................................................... 574

A Map Search System based on a Spatial Query Language
Yuanyuan Wang, Panote Siriayaya, Haruka Sakara, Yukiko Kawai, Keishi Tajima ................................. 578

FaiRank: An Interactive System to Explore Fairness of Ranking in Online Job Marketplaces
Ahmad Ghizzawi, Julien Marinescu, Shady Elbassuwni, Sihem Amer-Yahia, Gilles Bisson ...................... 582

MM-evolver: A Multi-model Evolution Management Tool
Michal Vavrek, Irena Holubova, Stefanie Schertzinger .......................................................... 586

Resense: Transparent Record and Replay of Sensor Data in the Internet of Things
Dimitrios Giouroukis, Julius Hülsmann, Janis von Bleichert, Morgan Geldenhuyx, Tim Stullich, Felipe Gutierrez, Jonas Traub, Kaustubh Beedkar, Volker Markl .......................................................... 590

Improving Named Entity Recognition using Deep Learning with Human in the Loop
Ticiana Coelho Da Silva, Régis Magalhães, José de Macêdo, David Araújo, Natanael Araújo, Vinicius de Melo, Pedro Olimpio, Paulo Rego, Aloísio Neto .......................................................... 594

Demonstrating data collections curation and exploration with CURARE
Genoveva Vargas-Solar, Gavin Kemp, Irving Hernandez Gallegos, Javier-Alfonso Espinosa-Oviedo, Catarina Ferrera da Silva, Parisa Ghodous .......................................................... 598

SparkER: Scaling Entity Resolution in Spark
Luca Gagliardelli, Giovanni Simonini, Domenico Beneventano, Sonia Bergamaschi ............................ 602

Exploring Interpretable Features for Large Time Series with SE4TeC
Jingwei Zuo, Karine Zeitouni, Yehia Taher .......................................................... 606

HASQL: a Method of Masking System Failures
Mark Hannum, Adi Zaimi, Mike Ponomarenko .......................................................... 610

Query-Driven Data Minimization with the DataEconomist
Peter K. Schwab, Julian Matschinske, Andreas M. Wahl, Klaus Meyer-Wegener ............................... 614

Short Papers

ITGC: Information-theoretic grid-based clustering
Sahar Behzadi Soheil, Hermann Hinterhauser, Claudia Plant .......................................................... 618

Adaptive Watermarks: A Concept Drift-based Approach for Predicting Event-Time Progress in Data Streams
Ahmed Awad, Jonas Traub, Sherif Sakr .......................................................... 622

Identifying Bias in Name Matching Tasks
Alexandros Karakasidis, Evaggelia Pitoura .......................................................... 626

Rock - Let the points roam to their clusters themselves
Anna Beer, Daniyal Kazempour, Thomas Seidl .......................................................... 630
Fast Trajectory Range Query with Discrete Frechet Distance
   Jiahao Zhang, Bo Tang, Yiu Man Lung .......................... 634

Repairing of Record Linkage: Turning Errors into Insight
   Quyen Bui-Nguyen, Qing Wang, Jingyu Shao, Dinusha Vatsalan ................. 638

Streaming HyperCube: A Massively Parallel Stream Join Algorithm
   Yuan Qiu, Serafeim Papadias, Ke Yi .................................. 642

Exploring Fairness of Ranking in Online Job Marketplaces
   Shady Elbassouini, Sihem Amer-Yahia, Christine El Atie, Ahmad Ghizzawi, Bilel Oualha .... 646

Snapshot Isolation for Transactional Stream Processing
   Philipp Götze, Kai-Uwe Sattler ........................................ 650

Recurrent Neural Networks for Dynamic User Intent Prediction in Human-Database Interaction
   Venkata Vamsikrishna Meduri, Kanchan Chowdhury, Mohamed Sarwat .............. 654

Optimal Algorithm for Profiling Dynamic Arrays with Finite Values
   Dingcheng Yang, Wenjian Yu, Junhui Deng, Shenghua Liu ........................ 658

Publishing Differentially Private Datasets via Stable Microaggregation
   Masooma Itfikhar, Qing Wang, Yu Lin ................................ 662

Range Query Processing for Monitoring Applications over Untrustworthy Clouds
   Hoang Van Tran, Tristan Allard, Laurent D'Orazio, Amr El Abbadi ............. 666

Towards Augmented Database Schemes by Discovery of Latent Visual Attributes
   Tomas Groupp, Ladislav Peska, Tomas Skopal ................................ 670

Workload-Driven and Robust Selection of Compression Schemes for Column Stores
   Martin Boissier, Max Jendruk .......................................... 674

CLRL: Feature Engineering for Cross-Language Record Linkage
   Öykü Özlem Çakal, Mohammad Mahdavi, Ziawasch Abedjan ..................... 678

Operational Stream Processing: Towards Scalable and Consistent Event-Driven Applications
   Asterios Katsifodimos, Marios Fragkoulis ................................ 682

Metropolis-Hastings Algorithms for Estimating Betweenness Centrality
   Mostafa Haghir Chehreghani, Tarek Abdessalem, Albert Bifet .................... 686

From Copernicus Big Data to Extreme Earth Analytics

Neuromorphic Hardware As Database Co-Processors: Potential and Limitations
   Thomas Heinis ................................................................. 694

Query Driven Data Labeling with Experts: Why Pay Twice?
   Eyal Dushkin, Shay Gershtein, Tova Milo, Slava Novgorodov ..................... 698

A Galaxy of Correlations
   Daniyal Kazempour, Lisa Krombholz, Peer Kröger, Thomas Seidl .................. 702

Insights into a running clockwork: On interactive process-aware clustering
   Daniyal Kazempour, Thomas Seidl ........................................ 706

Interpolation-friendly B-trees: Bridging the Gap Between Algorithmic and Learned Indexes
   Ali Hadian, Thomas Heinis ................................................. 710
SynthEdit: Format transformations by example using edit operations  
Alexander Teodor Bogatu, Alvaro A. A. Fernandes, Norman W. Paton, Nikolaos Konstantinou  
714

Triad Enumeration at Trillion-Scale Using a Single Commodity Machine  
Yudi Santoso, Alex Thomo, Venkatesh Srinivasan, Sean Chester  
718

Fast Truss Decomposition in Large-scale Probabilistic Graphs  
Fatemeh Esfahani, Jian Wu, Venkatesh Srinivasan, Alex Thomo, Kui Wu  
722

An Experimental Study on Network Immunization  
Alvis Logins, Panagiotis Karras  
726