Proceedings of the VLDB Endowment

Volume 14, No. 12 – July 2021

Editors in Chief:

Xin Luna Dong and Felix Naumann

Associate Editors:

Alon Halevy, Anastasia Ailamaki, Angela Bonifati, Arun Kumar, Ashraf Aboulnaga, Eugene Wu, Floris Geerts, Graham Cormode, Jeffrey Xu Yu, Jiannan Wang, Jingren Zhou, Jorge Arnulfo Quiané Ruiz, Juliana Freire, Jun Yang, Martin Theobald, Nesime Tatbul, Paolo Papotti, Rainer Gemulla, Stefan Manegold, Stratos Idreos, Surajit Chaudhuri, Xuemin Lin, Yi Chen, Yufei Tao, Zachary Ives, Zhifeng Bao

Publication Editors:

Thorsten Papenbrock and Hannes Mühleisen

PVLDB – Proceedings of the VLDB Endowment

Volume 14, No. 12, August 2021.

All papers published in this issue will be presented at the 47th International Conference on Very Large Data Bases, Copenhagen, Denmark, 2021.

Copyright 2021 VLDB Endowment

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/. For any use beyond those covered by this license, obtain permission by emailing info@vldb.org.

Volume 14, Number 12, August 2021

Pages i – vii and 2655 - 3252

ISSN 2150-8097

Available at: http://www.pvldb.org and https://dl.acm.org/journal/pvldb

TABLE OF CONTENTS

Front Matter

Copyright Notice i
Table of Contents ii
PVLDB Organization and Review Board – Vol. 14 ix
Demonstration Track Chairs and Reviewers – Vol. 14 xii
Industrial Track Chairs and Reviewers – Vol. 14 xiii
Editorial xiv

Demonstrations

KDV-Explorer: A Near Real-Time Kernel Density Visualization System for Spatial Analysis 2655
Tsz Nam Chan, Pak Lon Ip, Leong Hou U, Weng Hou Tong, Shivansh Mittal, Ye Li, Reynold Cheng

Refiner: A Reliable Incentive-Driven Federated Learning System Powered by Blockchain 2659
Zhebin Zhang, Dajie Dong, Yuhang Ma, Yilong Ying, Dawei Jiang, Ke Chen, Lidan Shou, Gang Chen

MultiCategory: Multi-model Query Processing Meets Category Theory and Functional Programming 2663
Valter Uotila, Jiaheng Lu, Dieter Gawlick, Zhen Hua Liu, Souripriya Das, Gregory Pogossiants

Cquirrel: Continuous Query Processing over Acyclic Relational Schemas 2667
Qichen Wang, Chaoqi Zhang, Danish Alsayed, Ke Yi, Bin Wu, Feifei Li, Chaoqun Zhan

DeFiHap: Detecting and Fixing HiveQL Anti-Patterns 2671
Yuetian Mao, Shuai Yuan, Nan Cui, Tianjiao Du, Beijun Shen, Yuting Chen

A Demonstration of KGLac: A Data Discovery and Enrichment Platform for Data Science 2675
Ahmed Helal, Mossad Helali, Khaled Ammar, Essam Mansour

Assessing the Existence of a Model in your Data with ADESIT 2679
Pierre Faure-Giovagnoli, Marie Le Guilly, Vasile-Marian Scuturici, Jean-Marc Petit

Path Advisor: A Multi-Functional Campus Map Tool for Shortest Path 2683
Yinzhao Yan, Raymond Chi-Wing Wong

Intermittent Human-in-the-Loop Model Selection using Cerebro: A Demonstration 2687
Liangde Li, Supun C Nakandala, Arun Kumar

Low-Latency Compilation of SQL Queries to Machine Code 2691
Henning Funke, Jens Teubner

Sound of Databases: Sonification of a Semantic Web Database Engine 2695
Sven Groppe, Rico Klinckenberg, Benjamin Warnke

HyMAC: A Hybrid Matrix Computation System 2699
Zihao Chen, Zhizhen Xu, Chen Xu, Juan Soto, Volker Markl, Weining Qian, Aoying Zhou

GraphScope: A One-Stop Large Graph Processing System 2703
Jingbo Xu, Zhanning Bai, Wenfei Fan, Longbin Lai, Xue Li, Zhao Li, Zhengping Qian, Lei Wang, Yanyan Wang, Wenyuan Yu, Kai Zeng, Jingren Zhou

Just Move It! Dynamic Parameter Allocation in Action 2707
Alexander Renz-Wieland, Tobias Drobisch, Zoi Kaoudi, Rainer Gemulla, Volker Markl

CBench: Demonstrating Comprehensive Evaluation of Question Answering Systems over Knowledge Graphs Through Deep Analysis of Benchmarks 2711
Abdelghny Orogat, Ahmed El-roby

PostCENN: PostgreSQL with Machine Learning Models for Cardinality Estimation 2715
Lucas Woltmann, Dominik Olwig, Claudio Hartmann, Dirk Habich, Wolfgang Lehner

DENOUNCER: Detection of Unfairness in Classifiers 2719
Jinyang Li, Yuval Moskovitch, H. V. Jagadish

A Demonstration of QARTA: An ML-based System for Accurate Map Services 2723
Sofiane Abbar, Rade Stanojevic, Mashaal Musleh, Mohamed M Elshrif, Mohamed Mokbel

TraNCE: Transforming Nested Collections Efficiently 2727
Jaclyn Smith, Michael Benedikt, Brandon Moore, Milos Nikolic

Debugging Missing Answers for Spark Queries over Nested Data with Breadcrumb 2731
Ralf Diestelkämper, Seokki Lee, Boris Glavic, Melanie Herschel

Demonstration of Panda: A Weakly Supervised Entity Matching System 2735
Renzhi Wu, Prem Sakala, Peng Li, Xu Chu, Yeye He

Automatic Data Acquisition for Deep Learning 2739
Jiabin Liu, Fu Zhu, Chengliang Chai, Yuyu Luo, Nan Tang

DBMind: A Self-Driving Platform in openGauss 2743
Xuanhe Zhou, Lianyuan Jin, Ji Sun, Xinyang Zhao, Xiang Yu, Shifu Li, Tianqing Wang, Kun Li, Luyang Liu

Demonstration of Dealer: An End-to-End Model Marketplace with Differential Privacy 2747
Jinfei Liu, Qiongqiong Lin, Jiayao Zhang, Kui Ren, Jian Lou, Junxu Liu, Li Xiong, Jian Pei, Jimeng Sun

Assassin: an Automatic claSSificAtion system baSed on algorithm SelectIoN 2751
Tianyu Mu, Hongzhi Wang, Shenghe Zheng, Shaoqing Zhang, Cheng Liang, Haoyun Tang

ATLANTIC: Making Database Differentially Private and Faster with Accuracy Guarantee 2755
Lei Cao, Dongqing Xiao, Yizhou Yan, Samuel Madden, Guoliang Li

Demonstration of Marius: Graph Embeddings with a Single Machine 2759
Anders Carlsson, Anze Xie, Jason M Mohoney, Roger Waleffe, Shanan Peters, Theodoros Rekatsinas, Shivaram Venkataraman

From Papers to Practice: The openclean Open-Source Data Cleaning Library 2763
Heiko Mueller, Sonia Castelo, Munaf A Qazi, Juliana Freire

Demonstration of Apperception: A Database Management System for Geospatial Video Data 2767
Vanessa Lin, Yongming Ge, Maureen Daum, Alvin Cheung, Brandon Haynes, Magdalena Balazinska

Automated energy consumption forecasting with EnForce 2771
Mary Karatzoglidi, Paraskevas Kerasiotis, Verena Kantere

RealGraph-Web: A Graph Analysis Platform on the Web 2775
Myung-Hwan Jang, Yong-Yeon Jo, Sang-Wook Kim

Interactive Demonstration of SQLCHECK 2779
Arthita Ghosh, Arpit Narechania, Visweswara Sai Prashanth Dintyala, Su Timurturkan, Joy Arulraj, Deven Bansod

T-Cove: An exposure tracing System based on Cleaning Wi-Fi Events on Organizational Premises 2783
Yiming Lin, Pramod Khargonekar, Sharad Mehrotra, Nalini Venkatasubramanian

Demonstration of Generating Explanations for Black-Box Algorithms Using Lewis..... 2787
Paul Y Wang, Sainyam Galhotra, Romila Pradhan, Babak Salimi

Auctus: A Dataset Search Engine for Data Discovery and Augmentation..... 2791
Sonia Castelo, Remi Rampin, Aécio Santos, Aline Bessa, Fernando Chirigati, Juliana Freire

A Demonstration of Relic: A System for REtrospective Lineage InferenCe of Data Workflows..... 2795
Mohammed Suhail Rehman, Silu Huang, Aaron J Elmore

SChain: A Scalable Consortium Blockchain Exploiting Intra- and Inter-Block Concurrency 2799
Zhihao Chen, Haizhen Zhuo, Quanqing Xu, Xiaodong Qi, Chengyu Zhu, Zhao Zhang, Cheqing Jin, Aoying Zhou, Ying Yan, Hui Zhang

EPICGen: An Experimental Platform for Indoor Congestion Generation and Forecasting 2803
Chrysovalantis Anastasiou, Constantinos Costa, Panos K. Chrysanthis, Cyrus Shahabi

Wikinegata: a Knowledge Base with Interesting Negative Statements..... 2807
Hiba Arnaout, Simon Razniewski, Gerhard Weikum, Jeff Z. Pan

Full Encryption: An end to end encryption mechanism in GaussDB..... 2811
Liang Guo, Jinwei Zhu, Jiayang Liu, Kun Cheng

DatAgent: The Imminent Age of Intelligent Data Assistants 2815
Antonis Mandamadiotis, Georgia Koutrika, Stavroula Eleftherakis, Apostolis Glenis, Dimitrios Skoutas, Yannis Stavrakas

DICE: Data Discovery by Example 2819
El Kindi Rezig, Anshul Bhandari, Anna Fariha, Benjamin Price, Allan Vanterpool, Vijay Gadepally, Michael Stonebraker

AnyOLAP: Analytical Processing of Arbitrary Data-Intensive Applications without ETL..... 2823
Felix Schuhknecht, Aaron Priesterroth, Justus Henneberg, Reza Salkhordeh

A Demonstration of the Exathlon Benchmarking Platform for Explainable Anomaly Detection..... 2827
Vincent Jacob, Fei Song, Arnaud Stiegler, Bijan Rad, Yanlei Diao, Nesime Tatbul

An Intermediate Representation for Hybrid Database and Machine Learning Workloads 2831
Amir Shaikhha, Maximilian Schleich, Dan Olteanu

How Divergent Is Your Data? 2835
Eliana Pastor, Andrew Gavgavian, Elena Baralis, Luca De Alfaro

An Extensible and Reusable Pipeline for Automated Utterance Paraphrases 2839
Auday Berro, Mohammad-ali Yaghub Zade Fard, Marcos Baez, Boualem Benatallah, Khalid Benabdeslem

Compliant Geo-distributed Data Processing in Action 2843
Kaustubh Beedkar, David Brekardin, Jorge Arnulfo Quiane Ruiz, Volker Markl

Query-Driven Video Event Processing for the Internet of Multimedia Things 2847
Piyush Yadav, Dhaval Salwala, Felipe Pontes, Praneet Dhingra, Edward Curry

A Demonstration of NoDA: Unified Access to NoSQL Stores 2851
Nikolaos Koutroumanis, Kousathanas Nikolaos, Christos Doulkeridis, Akrivi Vlachou

AutoExecutor: Predictive Parallelism for Spark SQL Queries 2855
Rathijit Sen, Abhishek Roy, Alekh Jindal, Rui Fang, Jeff Zheng, Xiaolei Liu, Ruiping Li

Catch a Blowfish Alive: A Demonstration of Policy-Aware Differential Privacy for Interactive Data Exploration 2859
Jiaxiang Liu, Karl Knopf, Yiqing Tan, Bolin Ding, Xi He

RONIN: Data Lake Exploration 2863
Paul Ouellette, Aidan Sciortino, Fatemeh Nargesian, Bahar Ghadiri Bashardoost, Erkang Zhu, Ken Pu, Renée J. Miller

SAND in Action: Subsequence Anomaly Detection for Streams 2867
Paul Boniol, John Paparrizos, Themis Palpanas, Michael Franklin

Valentine in Action: Matching Tabular Data at Scale 2871
Christos Koutras, Kyriakos Psarakis, George Siachamis, Andra Ionescu, Marios Fragkoulis, Angela Bonifati, Asterios Katsifodimos

GEDet: Detecting Erroneous Nodes with A Few Examples 2875
Sheng Guan, Hanchao Ma, Sutanay Choudhury, Yinghui Wu

Industrial and Applications

GraphScope: A Unified Engine for Big Graph Processing 2879
Wenfei Fan, Longbin Lai, Xue Li, Yong Li, Zhao Li, Zhengping Qian, Chao Tian, Lei Wang, Jingbo Xu, Youyang Yao, Qiang Yin, Wenyuan Yu, Kai Zeng, Kun Zhao, Jingren Zhou, Diwen Zhu, Rong Zhu

Davos: A System for Interactive Data-Driven Decision Making 2893
Zeyuan Shang, Emanuel Zgraggen, Benedetto Buratti, Philipp Eichmann, Navid Karimeddiny, Charlie Meyer, Wesley Runnels, Tim Kraska

Mixer: Efficiently Understanding and Retrieving Visual Content at Web-Scale 2906
Mengbai Xiao, An Qin, Yongwei Wu, Xinjie Huang, Xiaodong Zhang

Towards A Polyglot Framework for Factorized ML..... 2918
David A Justo, Shaoqing Yi, Lukas Stadler, Nadia Polikarpova, Arun Kumar

The End of Moore's Law and the Rise of The Data Processor..... 2932
Niv Dayan, Yuval Rochman, Iddo Naiss, Shmuel Dashevsky, Noam Rabinovich, Edward Bortnikov, Igal Maly, Ofer Frishman, Itai Ben Zion, Avraham, Moshe Twitto, Uri Beitler, Evgeni Ginzburg, Mark Mokryn

tf.data: A Machine Learning Data Processing Framework 2945
Derek Murray, Jiri Simsa, Ana Klimovic, Ihor Indyk

Not Black-Box Anymore! Enabling Analytics-Aware Optimizations in Teradata Vantage..... 2959
Mohamed Eltabakh, Anantha Subramanian, Awny Alomari, Mohammed Al-kateb, Sanjay Nair, Mahbub Hasan, Wellington Cabrera, Charles Zhang, Amit Kishore, Snigdha Prasad

Fangorn: Adaptive Execution Framework for Heterogeneous Workloads on Shared Clusters 2972
Yingda Chen, Jiamang Wang, Yifeng Lu, Ying Han, Zhiqiang Lv, Xuebin Min, Hua Cai, Wei Zhang, Haochuan Fan, Chao Li, Tao Guan, Wei Lin, Yangqing Jia, Jingren Zhou

Napa: Powering Scalable Data Warehousing with Robust Query Performance at Google..... 2986
Ankur Agiwal, Kevin Lai, Gokul Nath Babu Manoharan, Indrajit Roy, Jagan Sankaranarayanan, Hao Zhang, Tao Zou, Jim Chen, Min Chen, Ming Dai, Thanh Do, Haoyu Gao, Haoyan Geng, Raman Grover, Bo Huang, Yanlai Huang, Adam Li, Jianyi Liang, Tao Lin, Li Liu, Yao Liu, Xi Mao, Maya Meng, Prashant Mishra, Jay Patel, Rajesh Sr, Vijayshankar Raman, Sourashis Roy, Mayank Singh Shishodia, Tianhang Sun, Justin Tang, Jun Tatemura, Sagar Trehan, Ramkumar Vadali, Prasanna Venkatasubramanian, Joey Zhang, Kefei Zhang, Yupu Zhang, Zeleng Zhuang, Goetz Graefe, Divy Agrawal, Jeff Naughton, Sujata Kosalge, Hakan Hacigumus

The Art of Balance: A RateupDB Experience of Building a CPU/GPU Hybrid Database Product..... 2999
Rubao Lee, Minghong Zhou, Chi Li, Shenggang Hu, Jianping Teng, Dongyang Li, Xiaodong Zhang

RAMP-TAO: Layering Atomic Transactions on Facebook's Online TAO Data Store..... 3014
Audrey Cheng, Xiao Shi, Lu Pan, Anthony Simpson, Neil Wheaton, Shilpa Lawande, Nathan Bronson, Peter Bailis, Natacha Crooks, Ion Stoica

openGauss: An Autonomous Database System..... 3028
Guoliang Li, Xuanhe Zhou, Ji Sun, Xiang Yu, Yue Han, Lianyuan Jin, Wenbo Li, Tianqing Wang, Shifu Li

Hyperspace: The Indexing Subsystem of Azure Synapse..... 3043
Rahul Potharaju, Terry Kim, Eunjin Song, Wentao Wu, Lev Novik, Apoorve Dave, Pouria Pirzadeh, Andrew Fogarty, Gurleen Dhody, Jiying Li, Vidip Acharya, Sinduja Ramanujam, Nico Bruno, Cesar Galindo-legaria, Vivek Narasayya, Surajit Chaudhuri, Anil Nori, Tomas Talius, Raghu Ramakrishnan

SpeakNav: Voice-based Route Description Language Understanding for Template Driven Path Search..... 3056
Bolong Zheng, Lei Bi, Juan Cao, Hua Chai, Jun Fang, Lu Chen, Yunjun Gao, Xiaofang Zhou, Christian S Jensen

Railgun: managing large streaming windows under MAD requirements 3069
Ana Sofia Gomes, João Oliveirinha, Pedro Cardoso, Pedro Bizarro

Big Metadata: When Metadata is Big Data 3083
Pavan Edara, Mosha Pasumansky

Tanium Reveal: A Federated Search Engine for Querying
Unstructured File Data on Large Enterprise Networks 3096
Joshua F Stoddard, Adam Mustafa, Naveen Goela

Hazelcast Jet: Low-latency Stream Processing at the 99.99th Percentile 3110
Can Gencer, Marko Topolnik, Viliam Ďurina, Emin Demirci, Basri Kahveci, Ali Gürbüz, Jozsef Bartok, Grzegorz Gierlach, František Hartman, Ufuk Yilmaz, Ondřej Lukas, Mehmet Doğan, Mohamed Mandouh, Marios Fragkoulis, Asterios Katsifodimos

SparkCruise: Workload Optimization in Managed Spark Clusters at Microsoft 3122
Abhishek Roy, Alekh Jindal, Priyanka Gomatam, Xiating Ouyang, Ashit Gosalia, Nishkam Ravi, Swinky Mann, Prakhar Jain

Watermarks in Stream Processing Systems:
Semantics and Comparative Analysis of Apache Flink and Google Cloud Dataflow 3135
Edmon Begoli, Tyler Akidau, Slabva Chernyak, Fabian Hueske, Kathryn Knight, Kenneth Knowles, Daniel Mills, Dan Sotolongo

The Cosmos Big Data Platform at Microsoft:
Over a Decade of Progress and a Decade to Look Forward 3148
Conor Power, Hiren Patel, Alekh Jindal, Jyoti Leeka, Bob Jenkins, Michael Rys, Ed Triou, Dexin Zhu, Lucky Katahanas, Chakrapani Bhat Talapady, Josh Rowe, Fan Zhang, Rich Draves, Ivan Santa, Amrish Kumar

The Evolution of Amazon Redshift 3162
Ippokratis Pandis

Using VDMS to Index and Search 100M Images 3240
Luis Remis, Chaunte W Lacewell

Tutorials

On the Limits of Machine Knowledge: Completeness, Recall and Negation in
Web-scale Knowledge Bases 3175
Simon Razniewski, Hiba Arnaout, Shrestha Ghosh, Fabian Suchanek

Managing ML Pipelines: Feature Stores and the Coming Wave of Embedding Ecosystems 3178
Laurel Orr, Atindriyo Sanyal, Xiao Ling, Karan Goel, Megan Leszczynski

Data Augmentation for ML-driven Data Preparation and Integration 3182
Yuliang Li, Xiaolan Wang, Zhengjie Miao, Wang-Chiew Tan

Array DBMS: Past, Present, and (Near) Future 3186
Ramon Antonio Rodriges Zalipynis

Machine Learning for Databases 3190
Guoliang Li, Xuanhe Zhou, Lei Cao

Extending the Lifetime of NVM: Challenges and Opportunities 3194
Saeed Kargar, Faisal Nawab

New Trends in High-D Vector Similarity Search: AI-driven, Progressive, and Distributed 3198
Karima Echihabi, Themis Palpanas, Kostas Zoumpatianos

Machine Learning for Cloud Data Systems: The Promise, the Progress, and the Path Forward 3202
Alekh Jindal, Matteo Interlandi

Endowment Awards

It's not just Cookies and Tea 3206
Susan Davidson

Evolution of a Compiling Query Engine 3207
Thomas Neumann

Make Your Database System Dream of Electric Sheep: Towards Self-Driving Operation 3211
Andrew Pavlo, Matthew Butrovich, Lin Ma
Prashanth Menon, Wan Shen Lim, Dana Van Aken, William Zhang

Keynotes

Towards instance-optimized data systems 3222
Tim Kraska

Knowledge Graphs 2021: A Data Odyssey 3233
Gerhard Weikum

Panel

The future of data(base) education: Is the "cow book" dead? 3239
Zachary G. Ives, Rachel Pottinger, Arun Kumar, Johannes Gehrke, Jana Giceva

PVLDB ORGANIZATION AND REVIEW BOARD - Vol. 14

Editors in Chief of PVLDB
Xin Luna Dong (Facebook)
Felix Naumann (HPI, University of Potsdam)

Associate Editors of PVLDB
Ashraf Aboulnaga (Qatar Computing Research Institute, Hamad Bin Khalifa University)
Anastasia Ailamaki (EPFL)
Zhifeng Bao (RMIT University)
Angela Bonifati (Lyon 1 University)
Surajit Chaudhuri (Microsoft Research)
Yi Chen (New Jersey Institute of Technology)
Graham Cormode (University of Warwick)
Juliana Freire (New York University)
Floris Geerts (University of Antwerp)
Rainer Gemulla (University of Mannheim)
Alon Halevy (Facebook)
Stratos Idreos (Harvard University)
Zachary Ives (University of Pennsylvania)
Arun Kumar (UC San Diego)
Xuemin Lin (University of New South Wales)
Stefan Manegold (CWI, Leiden University)
Paolo Papotti (Eurecom)
Jorge Arnulfo Quiané Ruiz (Technical University of Berlin)
Yufei Tao (Chinese University of Hong Kong)
Nesime Tatbul (Intel Labs and MIT)
Martin Theobald (Université du Luxembourg)
Jiannan Wang (Simon Fraser University)
Eugene Wu (Columbia University)
Jun Yang (Duke University)
Jeffrey Xu Yu (The Chinese University of Hong Kong)
Jingren Zhou (Alibaba)

Publication Editors
Thorsten Papenbrock (Philipps University of Marburg)
Hannes Mühleisen (CWI)

PVLDB Managing Editor
Wolfgang Lehner (Dresden University of Technology)

PVLDB Advisory Committee
Divesh Srivastava (AT&T Labs-Research)
M. Tamer Özsu (University of Waterloo)
Juliana Freire (New York University)
Xin Luna Dong (Amazon)
Peter Boncz (CWI)
Lei Chen (Hong Kong University of Science and Technology)
Graham Cormode (University of Warwick)
Xiaofang Zhou (University of Queensland)
Magdalena Balazinska (University of Washington)
Fatma Ozcan (IBM Almaden)
Felix Naumann (HPI, University of Potsdam)
Peter Triantafillou (University of Warwick)

Panel Chairs
Susan Davidson, University of Pennsylvania
Theo Rekatsinas, University of Wisconsin, Madison

Tutorial Chairs
Gautam Das, University of Texas at Arlington
Ioana Manolescu, INRIA

Review Board

Abolfazl Asudeh (University of Illinois)
Ahmed Eldawy (University of California, Riverside)
Alan Fekete (University of Sydney)
Alekh Jindal (Microsoft)
Alex Ratner (University of Washington)
Altigran da Silva (Universidade Federal do Amazonas)
Anthony Tung (National University of Singapore)
Antonios Deligiannakis (Technical University of Crete)
Arijit Khan Nanyang (Technological University, Singapore)
Arnau Prat (Sparsity Technologies)
Ashwin Machanavajjhala (Duke University)
Asterios Katsifodimos (Technical University of Delft)
Avrilia Floratou (Microsoft)
Babak Salimi (University of Washington)
Badrish Chandramouli (Microsoft Research)
Beng Chin Ooi (National University of Singapore)
Bin Yang (Aalborg University)
Boris Glavic (Illinois Institute of Technology)
Byron Choi (Hong Kong Baptist University)
Carlos Scheidegger (University of Arizona)
Carsten Binnig (Technical University of Darmstadt)
Ce Zhang (ETH Zurich)
Chengfei Liu (Swinburne University of Technology)
Chengkai Li (University of Texas at Arlington)
Chris Jermaine (Rice University)
Christian Bizer (University of Mannheim)
Cong Yu (Google)
Daisy Zhe Wang (University of Florida)
Danica Porobic (Oracle)
Davide Mottin (Aarhus University)
Dimitris Papadias (Hong Kong University of Science and Technology)
Dong Deng (Rutgers University)
Eric Lo (Chinese University of Hong Kong)
Essam Mansour (Concordia University)
Fatma Ozcan (IBM Research)
Flip Korn (Google)
Florin Rusu (University of California, Merced)
Fotis Psallidas (Microsoft)
Francesco Bonchi (ISI Foundation)
Gao Cong (Nanyang Technological University)
George Fletcher (Technical University of Eindhoven)
Georgia Koutrika (Athena Research Center)
Hao Wei (Amazon)
Heiko Mueller (New York University)
Hong Cheng (Chinese University of Hong Kong)
Hongzhi Wang (Harbin Institute of Technology)
Hung Ngo (RelationalAI)
Immanuel Trummer (Cornell University)
Ingo Müller (ETH Zürich)
Jana Giceva (Technical University of Munich)
Jennie Rogers (Northwestern University)
Jeong-Hyon Hwang (University at Albany, State University of New York)
Jiaheng Lu (University of Helsinki)
Jianliang Xu (Hong Kong Baptist University)
Jianxin Li (Deakin University)
Jignesh Patel (University of Wisconsin)
Johann Gamper (Free University of Bozen-Bolzano)
Johannes Gehrke (Microsoft)
Jonas Traub (Technical University of Berlin)
Joy Arulraj (Georgia Tech)
Ju Fan (Renmin University of China)
K. Selçuk Candan (Arizona State University)
Kai Zeng (Alibaba)
Katja Hose (Aalborg University)
Ken Salem (University of Waterloo)
Kenneth A. Ross (Columbia University)
Khuzaima Daudjee (University of Waterloo)
Konstantinos Karanasos (Microsoft)
Laurel Orr (Stanford University)
Lei Chen (Hong Kong University of Science and Technology)
Lei Zou (Peking University)
Li Xiong (Emory University)
Lu Chen (Aalborg University)
Lu Qin (University of Technology Sydney)
Manasi Vartak (Verta)
Manos Athanassoulis (Boston University)
Manos Karpathiotakis (Facebook)
Marco Serafini (University of Massachusetts Amherst)
Marcos Antonio Vaz Salles (University of Copenhagen)
Mark Callaghan (MongoDB)
Markus Weimer (Microsoft)
Matei Zaharia (Stanford University, Databricks)
Matteo Interlandi (Microsoft)
Matthaios Olma (Microsoft Research)
Meihui Zhang Beijing (Institute of Technology)
Miao Qiao (University of Auckland)
Michael H. Böhlen (University of Zurich)
Michael Cafarella (University of Michigan)
Mirek Riedewald (Northeastern University)
Mohamed Mokbel (Qatar Computing Research Institute)
Mohamed Sarwat (Arizona State University)
Mohammad Sadoghi (University of California, Davis)
Mourad Ouzzani (Qatar Computing Research Institute, Hamad Bin Khalifa University)
Muhammad Aamir Cheema (Monash University)
Murat Demirbas (University at Buffalo, SUNY)
Nan Tang (Qatar Computing Research Institute, Hamad Bin Khalifa University)
Nick Koudas (University of Toronto)
Nikolaus Augsten (University of Salzburg)
Norman May (SAP)
Norman Paton (University of Manchester)
Odysseas Papapetrou (Technical University of Eindhoven)
Oliver A. Kennedy (University at Buffalo, SUNY)
Paolo Merialdo (Roma Tre University)
Paraschos Koutris (University of Wisconsin – Madison)
Peter Boncz (Centrum Wiskunde & Informatica)
Qin Zhang Indiana (University Bloomington)
Raja Appuswamy (Eurecom)
Ralf Schenkel (University of Trier)

Raul Castro Fernandez (University of Chicago)
Raymond Chi-Wing Wong (Hong Kong University of Science and Technology)
Reynold Cheng (The University of Hong Kong)
Reza Akbarinia (INRIA)
Ruoming Jin (Kent State University)
Ryan Johnson (Amazon Web Services)
S. Sudarshan (IIT Bombay)
Sanjay Krishnan (University of Chicago)
Saravanan Thirumuruganathan (Qatar Computing Research Institute, Hamad Bin Khalifa University)
Sebastian Schelter (University of Amsterdam)
Semih Salihoglu (University of Waterloo)
Senjuti Basu Roy (New Jersey Institute of Technology)
Shaoxu Song (Tsinghua University)
Shimin Chen (Chinese Academy of Sciences)
Sibo Wang (The Chinese University of Hong Kong)
Silu Huang (Microsoft Research)
Spyros Blanas (Ohio State University)
Srikanth Kandula (Microsoft Research)
Steffen Zeuch (German Research Centre for Artificial Intelligence - DFKI)
Stijn Vansummeren (Université libre de Bruxelles)
Sudeepa Roy (Duke University)
Sudip Roy (Google)
Tamer Özsu (University of Waterloo)
Themis Palpanas (University of Paris, French University Institute - IUF)
Tianzheng Wang (Simon Fraser University)
Tingjian Ge (University of Massachusetts, Lowell)
Thomas Heinis (Imperial College)
Thomas Neumann (Technical University of Munich)
Toon Calders (Universiteit Antwerpen)
Umar Farooq Minhas (Microsoft Research)
Viktor Leis (Friedrich Schiller University Jena)
Walid Aref (Purdue University)
Wei-Shinn Ku (Auburn University)
Weiren Yu (University of Warwick)
Wendy Hui Wang (Stevens Institute of Technology)
Wenjie Zhang (University of New South Wales)
Wolfgang Gatterbauer (Northeastern University)
Xi He (University of Waterloo)
Xiang Zhao (National University of Defence Technology)
Xiangyao Yu (University of Wisconsin – Madison)
Xiaokui Xiao (National University of Singapore)
Xiaolan Wang (Megagon Labs)
Xin Cao (University of New South Wales)
Xu Chu (Georgia Tech)
Yannis Velegrakis (Utrecht University)
Ye Yuan (Beijing Institute of Technology)
Yeye He (Microsoft Research)
Ying Zhang (University of Technology Sydney)
Yinghui Wu (Case Western Reserve University)
Yongjoo Park (University of Illinois at Urbana-Champaign)
Yongxin Tong (Beihang University)
Yu Yang (City University of Hong Kong)
Yuchen Li (Singapore Management University)
Yudian Zheng (Twitter)
Yunjun Gao (Zhejiang University)
Zechao Shang (University of Chicago)
Zhenjie Zhang (Singapore R&D, Yitu Technology Ltd.)
Zhewei Wei (Renmin University of China)
Ziawasch Abedjan (Technical University of Berlin)
Zoi Kaoudi (Technical University of Berlin)

DEMONSTRATION TRACK CHAIRS AND REVIEWERS - Vol. 14

Demonstration Track PC Chairs
Torsten Grust, University of Tuebingen, Germany
Guoliang Li, Tsinghua University, China
Yuanyuan Tian, Microsoft, USA

Demonstration Track Reviewers
Alvin Cheung, University of California, Berkeley, USA
Anna Fariha, University of Massachusetts Amherst, USA
Anja Gruenheid, Google Inc., USA
Avrilia Floratou, Microsoft, USA
Chengliang Chai, Tsinghua University, China
Da Yan, University of Alabama at Birmingham, USA
Dong Xie, Penn State University, USA
Evangelia Sitaridi, Amazon Web Services, USA
Felix Schuhknecht, Johannes Gutenberg-University Mainz, Germany
Gao Cong, Nanyang Technological Univesity, Singapore
George Fletcher, Eindhoven University of Technology, Netherlands
Hongzhi Wang, Harbin Institute of Technology, China
Hua Lu, Roskilde University, Denmark
Ingo Müller, ETH Zürich, Switzerland
Jana Giceva, TU Munich, Germany
Jia Yu, Washington State University, Germany
Jianguo Wang, Purdue University, USA
Jiannan Wang, Simon Fraser University, Canada
Ju Fan, Renmin University of China, China
Kai Zeng, Alibaba Group, China
Khuzaima Daudjee, University of Waterloo, Canada
Knut Stolze, IBM, Germany
Laure Berti-Equille, IRD, France
Lei Cao, MIT, USA
Lukas Rupprecht, IBM Research - Almaden, USA
Manuel Rigger, ETH Zurich, Switzerland
Mark Raasveldt, CWI, Netherlands
Michael Gubanov, Florida State University, USA
Ning Wang, Beijing Jiaotong University, China
Qun Chen, Northwestern Polytechnical University, China
Shuai Ma, Beihang University, China
Silu Huang, Microsoft, USA
Tarique Siddiqui, Microsoft Research, USA
Umar Farooq Minhas, Microsoft Research, USA
Vasilis Efthymiou, IBM Research - Almaden, USA
Wenjie Zhang, University of New South Wales, Australia
Wenlei Xie, Facebook, USA
Xiao Qin, IBM Research, USA
Xiaokui Xiao, National University of Singapore, Singapore
Yeye He, Microsoft Research, USA
Yingjun Wu, Amazon Web Services, USA
Yongxin Tong, Beihang University, China
Zhifeng Bao, RMIT University, Australia

INDUSTRIAL TRACK CHAIRS AND REVIEWERS - Vol. 14

Industrial Track PC Chairs
Castellanos, Teradata, USA
Chen Li, UC Irvine, USA
Feifei Li, Alibaba Group, China

Industrial Track Reviewers
Divy Agrawal UCSB, USA
Awny Alomari Teradata, USA
Roger Barga Amazon, USA
Paul Brown Teradata, USA
Yingyi Bu Google, USA
Jianjun Chen Bytedance, USA
Umeshwar Dayal Hitachi, USA
Bailu Ding Microsoft, USA
Dimitrios Georgakopoulos Swinsburne University of Technology, Australia
Georg Gottlob Oxford University, England
Yeye He Microsoft, USA
Meichun Hsu Oracle, USA
Yannis Ioannidis University of Athends, Greece
Jongik Kim Jeonbuk National University, Korea
Taewoo Kim Microsoft, USA
Alberto Laender Universidade Federal de Minas Gerais, Brazil
Jeff Lefevre UCSC, USA
Wolfgang Lehner Technische Universität Dresden, Germany
Guoliang Li Tsinghua University, China
Jinfeng Li Megagon, USA
Chunbin Lin Amazon, USA
Sergey Melnik Google, USA
Julio Navas SAP, USA
Beng Chin Ooi National University of Singapore, Singapore
Ippokratis Pandis Amazon, USA
Debjyoti Paul Facebook, USA
Cyrus Shahabi University of Southern California, USA
Kyuseok Shim Seoul National University, Korea
Leticia Tanca Polytechnic University of Milan, Italy
Jianguo Wang Purdue University, USA
Jiannan Wang Simon Fraiser University, Canada
Sheng Wang Alibaba Group, Singapore
Wei Wang UNSW, Australia
Steven Whang KAIST, Korea
Wolfram Wingerath Baqend, Germany
Chuan Xiao Osaka University, Japan
Kai Zeng Alibaba Group, China
Tieying Zhang Alibaba Group, USA
Roberto Zicari Goethe University Frankfurt, Germany

EDITORIAL

The Proceedings of the VLDB Endowment (PVLDB) provides a high-quality journal publication service to the data management research community. Each volume offers twelve monthly submission deadlines on the first day of each month and a quick, six-week reviewing cycle. This publication model was pioneered by PVLDB and combines a journal-style reviewing process, which includes a three-month revision cycle, with the agility and visibility provided by rapid on-line publication, and presentation at the annual VLDB conference.

PVLDB attracts many submissions spanning diverse data management topics, and the PVLDB reviewing process is implemented by a large team of dedicated researchers. The Review Board of PVLDB Volume 14 consists of 257 expert researchers, and reviewing is coordinated by 28 Associate Editors. Review board members provide timely (within a 3-week deadline) high-quality reviews, and participate actively in online discussions led by the Associate Editors for each paper. When needed, the Associate Editors together with the Editors-in-Chief solicit additional reviews from external experts. We give special thanks here to those additional reviewers who in most cases needed to complete their expert reviews on a very short notice.

Most of the accepted papers go through a revision process, which requires a second round of reviews after the authors have addressed an initial set of issues and concerns raised by the reviewers during the first round. Some papers are further accepted with shepherding, which means that one of the reviewers works with the authors to address a final set of comments.

This is the twelfth issue of the fourteenth volume of PVLDB, featuring 98 papers, all associated with the VLDB'21 conference: 56 demonstrations, 22 industrial track papers, eight tutorials, three invited papers for the VLDB Endowment awards, and one panel abstract.

The demonstration track at VLDB 2021 accepted 56 demonstrations out of a total of 130 submissions. The demos cover topics forming four big groups: Graph and other non-standard data types and applications, machine learning with and for database systems, aspects of query processing, and Data science, data cleaning and discovery.

The industrial track at VLDB 2021 accepted 22 papers from a total of 64 submissions. These works, reported by large and small companies from around the world, showcase the diversity of industrial research on machine learning systems, graph systems, big data systems, data warehousing, metadata management, stream processing, transaction processing, indexing, optimization, and other topics of interest to this conference.

This issue also includes descriptions of the eight tutorials presented at the VLDB 2021 conference. Those tutorials covered timely and exciting topics: ML for databases and data management for ML pipelines, knowledge graphs, array databases, similarity search, and non-volatile memory.

VLDB 2021 have 12 invited talks, including three VLDB Endowment award talks, two keynotes, six invited talks in the Scalable Data Science Invited Talk Series, and one invited talk for the Industry Track. The 2021 VLDB Early Career Research Contribution Award this year went to Prof. Andy Pavlo for contributions work on self-tuning databases. The VLDB Women in Database Research Award went to Prof. Susan Davidson. Finally, the 2021 VLDB Test of Time Award went to Prof. Thomas Neumann for his PVLDB Vol. 4 paper entitled "Efficiently compiling efficient query plans for modern hardware". For the two keynotes, Prof. Gerhard Weikum spoke on "Knowledge Graphs 2021: a Data Odyssey", and Prof. Tim Kraska talked about "Towards Instance-Optimized Data Systems." This issue includes extended abstracts for the above talks.

The Scalable Data Science invited talk series is newly introduced this year. We have invited six well-established researchers from academia and industry across continents talking about their data science practices. They cover topics

including online ML, production-friendly ML, intelligent ML applications, graph mining, and data summarization. In addition, the Industry keynote speaker, Ippokratis Pandis, will talk about the evolution of Amazon Redshift.

Finally, the VLDB 2021 conference also included a panel entitled "The Future of Data(base) Education: Is the 'Cow Book' Dead?", and an abstract for the panel is included in this issue.

These papers will be presented at the 47th International Conference on Very Large Data Bases (VLDB 2021) between Aug 16 and Aug 20. The conference will be in a hybrid mode, with both participants coming in person to Copenhagen, Denmark, and online audience. We hope that the readers will find the selected papers engaging, and thought-provoking. We also hope that the selected papers will provide valuable insights and inspire novel systems contributions and follow-up research.

Xin Luna Dong and Felix Naumann
PVLDB Volume 14 Editors in Chief