



# Proceedings of the VLDB Endowment

Volume 15, No. 12 – August 2022

Editors in Chief:

**Fatma Özcan, Juliana Freire and Xuemin Lin**

Associate Editors:

**Arun Kumar, Azza Abouzied, Beng Chin Ooi, Boris Glavic, Dan Suciu,  
Divyakant Agrawal, Eugene Wu, Georgia Koutrika, Ioana Manolescu,  
Jeffrey Xu Yu, Julia Stoyanovich, Jun Yang, K. Selçuk Candan,  
Khuzaima Daudjee, Laure Berti-Equille, Lei Chen, Mohamed Mokbel,  
Neoklis Polyzotis, Paolo Papotti, Peter Boncz, Sebastian Schelter,  
Sourav S Bhowmick, Surajit Chaudhuri, Themis Palpanas, Vanessa Braganholo,  
Viktor Leis, Wang-Chiew Tan, Wenjie Zhang, Wook-Shin Han, Xiaofang Zhou**

Publication Editors:

**Lijun Chang and Xin Cao**

PVLDB – Proceedings of the VLDB Endowment

Volume 15, No. 12, August 2022.

All papers published in this issue will be presented at the 48th International Conference on Very Large Data Bases, Sydney, Australia, 2022.

## **Copyright 2022 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](mailto:info@vldb.org).

Volume 15, Number 12, August 2022

Pages i – xiv and 3277 - 3827

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. 15 .....	viii
Industrial Track Chairs and Reviewers – Vol. 15 .....	xi
Demonstration Track Chairs and Reviewers – Vol. 15 .....	xii

### Industrial Papers

Hardware Acceleration of Compression and Encryption in SAP HANA.....	3277
<i>Monica Chiosa, Fabio Maschi, Ingo Müller, Gustavo Alonso, Norman May</i>	
Frost: A Platform for Benchmarking and Exploring Data Matching Results.....	3292
<i>Martin Graf, Lukas Laskowski, Florian Papsdorf, Florian Sold, Roland Gremmelspacher, Felix Naumann, Fabian Panse</i>	
ByteGraph: A High-Performance Distributed Graph Database in ByteDance .....	3306
<i>Changji Li, Hongzhi Chen, Shuai Zhang, Yingqian Hu, Chao Chen, Zhenjie Zhang, Meng Li, Xiangchen Li, Dongqing Han, Xiaohui Chen, Xudong Wang, Huiming Zhu, Xuwei Fu, Tingwei Wu, Hongfei Tan, Hengtian Ding, Mengjin Liu, Kangcheng Wang, Ting Ye, Lei Li, Xin Li, Yu Wang, Chenguang Zheng, Hao Yang, James Cheng</i>	
CDI-E: An Elastic Cloud Service for Data Engineering.....	3319
<i>Prakash C Das, Shivangi Srivastava, Valentin Moskovich, Anmol Chaturvedi, Anant Mittal, Yongqin Xiao, Mosharaf Chowdhury</i>	
Operon: An Encrypted Database for Ownership-Preserving Data Management .....	3332
<i>Sheng Wang, Yiran Li, Huorong Li, Feifei Li, Chengjin Tian, Le Su, Yanshan Zhang, Yubing Ma, Lie Yan, Yuanyuan Sun, Xuntao Cheng, Xiaolong Xie, Yu Zou</i>	
Tair-PMem: a Fully Durable Non-Volatile Memory Database.....	3346
<i>Caixin Gong, Chengjin Tian, Zhengheng Wang, Sheng Wang, Xiyu Wang, Qiulei Fu, Wu Qin, Qian Long, Rui Chen, Jiang Qi, Ruo Wang, Guoyun Zhu, Chenghu Yang, Wei Zhang, Feifei Li</i>	
Trie memtables in Cassandra.....	3359
<i>Branimir Lambov</i>	
Velox: Meta’s Unified Execution Engine.....	3372
<i>Pedro Pedreira, Orri Erling, Maria Basmanova, Kevin Wilfong, Laith S Sakka, Krishna Pai, Wei He, Biswapesh Chattopadhyay</i>	
OceanBase: A 707 Million tpmC Distributed Relational Database System.....	3385
<i>Zhenkun Yang, Chuanhui Yang, Fusheng Han, Mingqiang Zhuang, Bing Yang, Zhifeng Yang, Xiaojun Cheng, Yuzhong Zhao, Wenhui Shi, Huafeng Xi, Huang Yu, Bin Liu, Yi Pan, Boxue Yin, Junquan Chen, Quanqing Xu</i>	
VRE: A Versatile, Robust, and Economical Trajectory Data System.....	3398
<i>Hai Lan, Jiong Xie, Zhifeng Bao, Feifei Li, Wei Tian, Fang Wang, Sheng Wang, Ailin Zhang</i>	

ByteHTAP: ByteDance’s HTAP System with High Data Freshness and Strong Data Consistency.....	3411
<i>Jianjun Chen, Yonghua Ding, Ye Liu, Fangshi Li, Li Zhang, Mingyi Zhang, Kui Wei, Lixun Cao, Dan Zou, Yang Liu, Lei Zhang, Rui Shi, Wei Ding, Kai Wu, Shangyu Luo, Jason Sun, Yuming Liang</i>	
Beaconnect: Continuous Web Performance A/B Testing at Scale.....	3425
<i>Wolfram Wingerath, Benjamin Wollmer, Markus Bestehorn, Stephan Succo, Sophie Ferrlein, Florian Bücklers, Jörn Domnik, Fabian Panse, Erik Witt, Anil Sener, Felix Gessert, Norbert Ritter</i>	
CloudJump: Optimizing Cloud Databases for Cloud Storages .....	3432
<i>Zongzhi Chen, Xinjun Yang, Feifei Li, Xuntao Cheng, Qingda Hu, Zheyu Miao, Rongbiao Xie, Xiaofei Wu, Kang Wang, Zhao Song, Haiqing Sun, Zechao Zhuang, Yuming Yang, Jie Xu, Liang Yin, Wenchao Zhou, Sheng Wang</i>	
DyHealth: Making Neural Networks Dynamic for Effective Healthcare Analytics .....	3445
<i>Kaiping Zheng, Shaofeng Cai, Horng-ruey Chua, Melanie Herschel, Meihui Zhang, Beng Chin Ooi</i>	
Blueprint: a constraint-solving approach for document extraction .....	3459
<i>Andrey Mishchenko, Dominique Danco, Abhilash Jindal, Adrian Blue</i>	
TencentCLS: The Cloud Log Service with High Query Performances .....	3472
<i>Muzhi Yu, Zhaoxiang Lin, Jinan Sun, Runyun Zhou, Guoqiang Jiang, Hua Huang, Shikun Zhang</i>	
Ganos: A Multidimensional, Dynamic, and Scene-Oriented Cloud-Native Spatial Database Engine ..	3483
<i>Jiong Xie, Zhen Chen, Jianwei Liu, Fang Wang, Feifei Li, Zhida Chen, Yinpei Liu, Songlu Cai, Zhenhua Fan, Fei Xiao, Yue Chen</i>	
Magma: A high data density storage engine used in Couchbase.....	3496
<i>Sarath Lakshman, Apaar Gupta, Rohan Suri, Scott D Lashley, John Liang, Srinath Duvuru, Ravi Mayuram</i>	
Doppler: Automated SKU Recommendation in Migrating SQL Workloads to the Cloud.....	3509
<i>Joyce Cahoon, Wenjing Wang, Yiwen Zhu, Katherine Lin, Sean Liu, Raymond Truong, Neetu Singh, Chengcheng Wan, Alexandra M Ciortea, Sreraman Narasimhan, Subru Krishnan</i>	
Meta's Next-generation Realtime Monitoring and Analytics Platform .....	3522
<i>Stavros Harizopoulos, Taylor Hopper, Morton Mo, Shyam Sundar Chandrasekaran, Tongguang Chen, Yan Cui, Nandini Ganesh, Gary Helmling, Hieu Pham, Sebastian Wong</i>	
SQLite: Past, Present, and Future .....	3535
<i>Kevin P Gaffney, Martin Prammer, Laurence C Brasfield, Richard Hipp, Dan R Kennedy, Jignesh Patel</i>	
Manu: A Cloud Native Vector Database Management System .....	3548
<i>Rentong Guo, Xiaofan Luan, Long Xiang, Xiao Yan, Xiaomeng Yi, Jigao Luo, Qianya Cheng, Weizhi Xu, Jiarui Luo, Frank Liu, Zhenshan Cao, Yanliang Qiao, Ting Wang, Bo Tang, Charles Xie</i>	

## Demonstrations

Automated Relational Data Explanation using External Semantic Knowledge.....	3562
<i>Sainyam Galhotra, Udayan Khurana</i>	
Kelpie: an Explainability Framework for Embedding-based Link Prediction Models .....	3566
<i>Andrea Rossi, Donatella Firmani, Paolo Merialdo, Tommaso Teofili</i>	

OREO: Detection of Cherry-picked Generalizations .....	3570
<i>Yin Lin, Brit Youngmann, Yuval Moskovitch, H. V. Jagadish, Tova Milo</i>	
DuckDB-Wasm: Fast Analytical Processing for the Web .....	3574
<i>André Kohn, Dominik Moritz, Mark Raasveldt, Hannes Mühleisen, Thomas Neumann</i>	
EasyDR: A Human-in-the-loop Error Detection&Repair Platform for Holistic Table Cleaning .....	3578
<i>Yihai Xi, Ning Wang, Xinyu Chen, Yiyi Zhang, Zilong Wang, Zhihong Xu, Yue Wang</i>	
Hu-Fu: A Data Federation System for Secure Spatial Queries.....	3582
<i>Xuchen Pan, Yongxin Tong, Chunbo Xue, Zimu Zhou, Junping Du, Yuxiang Zeng, Yexuan Shi, Xiaofei Zhang, Lei Chen, Yi Xu, Ke Xu, Weifeng Lv</i>	
Demonstrating CAT: Synthesizing Data-Aware Conversational Agents for Transactional Databases	3586
<i>Marius Gassen, Benjamin Hättasch, Benjamin Hilprecht, Nadja Geisler, Alexander Fraser, Carsten Binnig</i>	
EDA4SUM: Guided Exploration of Data Summaries .....	3590
<i>Aurélien Personnaz, Brit Youngmann, Sihem Amer-yahia</i>	
CaJaDE: Explaining Query Results by Augmenting Provenance with Context .....	3594
<i>Chenjie Li, Juseung Lee, Zhengjie Miao, Boris Glavic, Sudeepa Roy</i>	
Share the Tensor Tea: How Databases can Leverage the Machine Learning Ecosystem .....	3598
<i>Yuki Asada, Victor Fu, Apurva Gandhi, Advitya Gemawat, Lihao Zhang, Vivek Gupta, Ehi Nosakhare, Dalitso Banda, Rathijit Sen, Matteo Interlandi</i>	
MOCHA: A Tool for Visualizing Impact of Operator Choices in Query Execution Plans for Database Education .....	3602
<i>Jess Tan, Desmond Yeoh, Rachael Neoh, Huey Eng Chua, Sourav S Bhowmick</i>	
LIBKDV: A Versatile Kernel Density Visualization Library for Geospatial Analytics.....	3606
<i>Tsz Nam Chan, Pak Lon Ip, Kaiyan Zhao, Leong Hou U, Byron Choi, Jianliang Xu</i>	
A Demonstration of Multi-Region CockroachDB .....	3610
<i>Arul Ajmani, Aayush Shah, Alexander Shraer, Adam Storm, Rebecca Taft, Oliver Tan, Nathan Vanbenschoten</i>	
DPDS: Assisting Data Science with Data Provenance.....	3614
<i>Adriane Chapman, Luca Lauro, Paolo Missier, Riccardo Torlone</i>	
POEM: Pattern-Oriented Explanations of CNN Models .....	3618
<i>Vargha Dadvar, Lukasz Golab, Divesh Srivastava</i>	
WebArrayDB: A Geospatial Array DBMS in Your Web Browser.....	3622
<i>Ramon Antonio Rodrigues Zalipynis, Nikita A Terlych</i>	
AutoDI: Towards an Automatic Plan Regression Analysis.....	3626
<i>Hai Lan, Yuanjia Zhang, Zhifeng Bao, Yu Dong, Dongxu Huang, Liu Tang, Jian Zhang</i>	
PHOCUS: Efficiently Archiving Photos .....	3630
<i>Susan B Davidson, Shay Gershtein, Tova Milo, Slava Novgorodov, May Shoshan</i>	

VINCENT: Towards Efficient Exploratory Subgraph Search in Graph Databases .....	3634
<i>Kai Huang, Qingqing Ye, Jing Zhao, Xi Zhao, Haibo Hu, Xiaofang Zhou</i>	
ActivePDB: Active Probabilistic Databases .....	3638
<i>Osnat Drien, Matanya Freiman, Yael Amsterdamer</i>	
CERTEM: Explaining and Debugging Black-box Entity Resolution Systems with CERTA.....	3642
<i>Tommaso Teofili, Donatella Firmani, Nick Koudas, Paolo Merialdo, Divesh Srivastava</i>	
Satellite Image Search in AgoraEO .....	3646
<i>Ahmet Kerem Aksoy, Pavel Dushev, Eleni Tzirita Zacharitou, Holmer Hensen, Marcela Charfuelan, Jorge Arnulfo Quiane Ruiz, Begum Demir, Volker Markl</i>	
SENSOR: Data-driven Construction of Sketch-based Visual Query Interfaces for Time Series Data .	3650
<i>Li Yan, Nerissa Xu, Guozhong Li, Sourav S Bhowmick, Byron Choi, Jianliang Xu</i>	
DiscoPG: Property Graph Schema Discovery and Exploration .....	3654
<i>Angela Bonifati, Stefania G. Dumbrava, Emile Martinez, Fatemeh Ghasemi, Malo Jaffré, Pacome Luton, Thomas Pickles</i>	
SA-Q: Observing, Evaluating, and Enhancing the Quality of the Results of Sentiment Analysis Tools .....	3658
<i>Wissam Mammam Kouadri, Salima Benbernou, Mourad Ouziri, Themis Palpanas, Iheb Benamor</i>	
SmartBench: Demonstrating Automatic Generation of Comprehensive Benchmarks for Question Answering Over Knowledge Graphs.....	3662
<i>Abdelghny Orogat, Ahmed El-robby</i>	
DADER: Hands-Off Entity Resolution with Domain Adaptation.....	3666
<i>Jianhong Tu, Xiaoyue Han, Ju Fan, Nan Tang, Chengliang Chai, Guoliang Li, Xiaoyong Du</i>	
Sigma Workbook: A Spreadsheet for Cloud Data Warehouses.....	3670
<i>James L Gale, Max Seiden, Deepanshu Utkarsh, Jason Frantz, Rob Woollen, Cagatay Demiralp</i>	
ReMac: A Matrix Computation System with Redundancy Elimination.....	3674
<i>Zihao Chen, Zhizhen Xu, Baokun Han, Chen Xu, Weining Qian, Aoying Zhou</i>	
TimeEval: A Benchmarking Toolkit for Time Series Anomaly Detection Algorithms.....	3678
<i>Phillip Wenig, Sebastian Schmidl, Thorsten Papenbrock</i>	
DBMS Annihilator: A High-Performance Database Workload Generator in Action.....	3682
<i>Alberto Lerner, Matthias Jasný, Theo Jepsen, Carsten Binnig, Philippe Cudre-mauroux</i>	
FedTSC: A Secure Federated Learning System for Interpretable Time Series Classification .....	3686
<i>Zhiyu Liang, Hongzhi Wang</i>	
AMRAS: A Visual Analysis System for Spatial Crowdsourcing.....	3690
<i>Qingshun Wu, Yafei Li, Huiling Li, Di Zhang, Guanglei Zhu</i>	
SparkCAD: Caching Anomalies Detector for Spark Applications .....	3694
<i>Hani Al-sayeh, Muhammad Attahir Jibril, Muhammad Waleed Bin Saeed, Kai-uwe Sattler</i>	

AvantGraph Query Processing Engine.....	3698
<i>Wilco Van Leeuwen, Thomas Mulder, Bram Van De Wall, George Fletcher, Nikolay Yakovets</i>	
Theseus: Navigating the Labyrinth of Time-Series Anomaly Detection .....	3702
<i>Paul Boniol, John Paparrizos, Yuhao Kang, Themis Palpanas, Ruey S. Tsay, Aaron J Elmore, Michael Franklin</i>	
A Demonstration of AutoOD: A Self-tuning Anomaly Detection System.....	3706
<i>Dennis M Hofmann, Peter Vannostrand, Huayi Zhang, Yizhou Yan, Lei Cao, Samuel Madden, Elke A Rundensteiner</i>	
Pipemizer: An Optimizer for Analytics Data Pipelines.....	3710
<i>Sunny Gakhar, Joyce Cahoon, Wangchao Le, Xiangnan Li, Kaushik Ravichandran, Hiren Patel, Marc Friedman, Brandon Haynes, Shi Qiao, Alekh Jindal, Jyoti Leeka</i>	
DORIAN in action: Assisted Design of Data Science Pipelines.....	3714
<i>Sergey Redyuk, Zoi Kaoudi, Sebastian Schelter, Volker Markl</i>	
WebMILE: Democratizing Network Representation Learning at Scale.....	3718
<i>Yuntian He, Yue Zhang, Saket Gurukar, Srinivasan Parthasarathy</i>	
Demonstrating Quest: A Query-Driven Framework to Explain Classification Models on Tabular Data .....	3722
<i>Nadja Geisler, Benjamin Hättasch, Carsten Binnig</i>	
IsoBugView: Interactively Debugging Isolation Bugs in Database Applications .....	3726
<i>Drew Ripberger, Yifan Gan, Xueyuan Ren, Spyros Blanas, Yang Wang</i>	
YeSQL: Rich User-Defined Functions without the Overhead .....	3730
<i>Yannis E Fofoulas, Alkis Simitsis, Yannis Ioannidis</i>	
Demonstration of Accelerating Machine Learning Inference Queries with Correlative Proxy Models	3734
<i>Zhihui Yang, Yicong Huang, Zuozhi Wang, Feng Gao, Yao Lu, Chen Li, X. Sean Wang</i>	
Demonstration of Collaborative and Interactive Workflow-Based Data Analytics in Texera.....	3738
<i>Xiaozhen Liu, Zuozhi Wang, Shengquan Ni, Sadeem Alsudais, Yicong Huang, Avinash Kumar, Chen Li</i>	
SimDB in Action: Road Traffic Simulations Completely Inside Array DBMS.....	3742
<i>Ramon Antonio Rodrigues Zalipynis</i>	

## Tutorials

Transformers for Tabular Data Representation: A Tutorial on Models and Applications.....	3746
<i>Gilbert Badaro, Paolo Papotti</i>	
Polyglot Data Management: State of the Art & Open Challenges.....	3750
<i>Felix Kiehn, Mareike Schmidt, Daniel Glake, Fabian Panse, Wolfram Wingerath, Benjamin Wollmer, Martin Poppinga, Norbert Ritter</i>	
Machine Programming: Turning Data into Programmer Productivity .....	3754
<i>Abdul Wasay, Nesime Tatbul, Justin Gottschlich</i>	

Cloud Databases: New Techniques, Challenges, and Opportunities .....	3758
<i>Guoliang Li, Haowen Dong, Chao Zhang</i>	
Modern Techniques for Querying Graph-Structured Relations: Foundations, System Implementations, and Open Challenges .....	3762
<i>Amine Mhedhbi, Semih Salihoglu</i>	
Densest Subgraph Discovery on Large Graphs: Applications, Challenges, and Techniques .....	3766
<i>Yixiang Fang, Wensheng Luo, Chenhao Ma</i>	
From BERT to GPT-3 Codex: Harnessing the Potential of Very Large Language Models for Data Management.....	3770
<i>Immanuel Trummer</i>	
The Past, Present and Future of Indexing on Persistent Memory.....	3774
<i>Kaisong Huang, Yuliang He, Tianzheng Wang</i>	
Unified Data Analytics: State-of-the-art and Open Problems .....	3778
<i>Zoi Kaoudi, Jorge Arnulfo Quiane Ruiz</i>	

## Keynotes

Big Graphs: Challenges and Opportunities .....	3782
<i>Wenfei Fan</i>	
Towards AI-Powered Data-Driven Education .....	3798
<i>Sihem Amer-yahia</i>	

## Endowment Awards

Heterogeneous Information Networks: the Past, the Present, and the Future.....	3807
<i>Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, Tianyi Wu</i>	
Toward Interpretable and Actionable Data Analysis with Explanations and Causality.....	3812
<i>Sudeepa Roy</i>	
Reflections On My Data Management Research Journey (VLDB Women in Database Research Award Talk) .....	3821
<i>Fatma Özcan</i>	

## Panel

Panel: Startups Founded by Database Researchers.....	3823
<i>C. Mohan</i>	
Cloud Data Systems: What are the Opportunities for the Database Research Community? .....	3826
<i>Magdalena Balazinska, Surajit Chaudhuri, Anhai Doan, Joseph M. Hellerstein, Hanuma Kodavalla, Ippokratis Pandis, Matei Zaharia</i>	



## **PVLDB ORGANIZATION AND REVIEW BOARD - Vol. 15**

### **Editors in Chief of PVLDB**

Fatma Ozcan (Google)  
Juliana Freire (New York University)  
Xuemin Lin (University of New South Wales)

### **Associate Editors of PVLDB**

Arun Kumar (University of California, San Diego)  
Azza Abouzied (NYU Abu Dhabi)  
Beng Chin Ooi (NUS)  
Boris Glavic (Illinois Institute of Technology)  
Dan Suciu (University of Washington)  
Divyakant Agrawal (University of California, Santa Barbara)  
Eugene Wu (Columbia University)  
Georgia Koutrika (ATHENA)  
Ioana Manolescu (INRIA and Institut Polytechnique de Paris)  
Jeffrey Xu Yu (Chinese University of Hong Kong)  
Julia Stoyanovich (New York University)  
Jun Yang (Duke University)  
K. Seçuk Candan (Arizona State University)  
Khuzaima Daudjee (University of Waterloo)  
Laks Lakshmanan (The University of British Columbia)  
Laure Berti-Equille (IRD)  
Lei Chen (Hong Kong University of Science and Technology)  
Mohamed Mokbel (University of Minnesota, Twin Cities)  
Neoklis Polyzotis (Google)  
Paolo Papotti  
Peter Boncz (CWI)  
Sebastian Schelter (University of Amsterdam)  
Sharad Mehrotra (U.C. Irvine)  
Sourav S Bhowmick (Nanyang Technological University)

Surajit Chaudhuri (Microsoft Research)  
Themis Palpanas (University of Paris)  
Vanessa Braganholo (Fluminense Federal University)  
Viktor Leis (Friedrich Schiller University Jena)  
Wang-Chiew Tan (Megagon Labs)  
Wenjie Zhang (University of New South Wales)  
Wook-Shin Han (POSTECH)  
Xiaofang Zhou (Hong Kong University of Science and Technology)

### **Publication Editors**

Lijun Chang (University of Sydney)  
Xin Cao (University of New South Wales)

### **PVLDB Managing Editor**

Wolfgang Lehner (Dresden University of Technology)

### **PVLDB Advisory Committee**

Felix Naumann (HPI)  
Juliana Freire (New York University)  
Xuemin Lin (U of New South Wales)  
Georgia Koutrika (Athena Research Center)  
Jun Yang (Duke University)  
Vanessa Braganholo (Universidade Federal Fluminense)  
Sourav S Bhowmick (Nanyang Technological University)  
Chris Jermaine (Rice University)  
Peter Triantafillou (University of Warwick)  
Xin Luna Dong (Facebook)  
Fatma Ozcan (Google)  
Lei Chen (Hong Kong University of S&T)  
Graham Cormode (University of Warwick)  
Divesh Srivastava (AT&T Labs-Research)  
Wolfgang Lehner (TU Dresden)

## Review Board

Abolfazl Asudeh (University of Michigan)  
Aécio Santos (New York University)  
Ahmed Eldawy (University of California, Riverside)  
Alexander Hall (RelationalAI)  
Alexander J Ratner (University of Washington)  
Aline Bessa (New York University)  
Alkis Simitsis (Athena Research Center)  
Altigran da Silva (Universidade Federal do Amazonas)  
AnHai Doan (University of Wisconsin-Madison)  
Anna Fariha (Microsoft)  
Anton Dignös (Free University of Bozen-Bolzano)  
Antonio Cavalcante Araujo Neto (University of Alberta)  
Arijit Khan (Nanyang Technological University)  
Arvind Arasu (Microsoft)  
Babak Salimi (University of California, San Diego)  
Bailu Ding (Microsoft Research)  
Bertram Ludaescher (University of Illinois)  
Bolong Zheng (Huazhong University of Science and Technology)  
Brandon Haynes (Gray Systems Lab, Microsoft)  
Byron Choi (Hong Kong Baptist University)  
Carlo Curino (Microsoft -- GSL)  
Carlos Scheidegger (The University of Arizona)  
Carsten Binnig (TU Darmstadt)  
Ce Zhang (ETH)  
Cheng Long (Nanyang Technological University)  
Chengfei Liu (Swinburne University of Technology)  
Chuan Lei (Instacart)  
Chunbin Lin (Amazon AWS)  
Curtis Dyreson (Utah State University)  
Dan Kifer (Pennsylvania State University)  
Dana M Van Aken (Carnegie Mellon University)  
Daniel Deutch (Tel Aviv University)  
Daniel Oliveira (UFF, Brazil)  
David Koop (Northern Illinois University)  
Davide Mottin (Aarhus University)  
Dong Xie (Penn State University)  
Eduardo Ogasawara (CEFET-RJ)  
Eleni Tzirita Zacharitou (TU Berlin)  
Fabio Porto (LNCC)  
Faisal Nawab (University of California at Irvine)  
Fan Zhang (Guangzhou University)  
Fatemeh Nargesian (University of Rochester)  
Fei Chiang (McMaster University)  
Florin Rusu (UC Merced)  
Floris Geerts (University of Antwerp)  
Fotis Psallidas (Microsoft)  
George Fletcher (Eindhoven University of Technology)  
George Papadakis (University of Athens)  
Gerhard Weikum (Max-Planck-Institut für Informatik)  
Germain Forestier (University of Haute Alsace)  
Guoliang Li (Tsinghua University)  
Haipeng Dai (Nanjing University)  
Harish Doraiswamy (Microsoft Research India)  
Heiko Mueller (DeepReason.ai)  
Herodotos Herodotou (Cyprus University of Technology)

Holger Pirk (Imperial College)  
Hongzhi Yin (The University of Queensland)  
Huiping Cao (New Mexico State University)  
Immanuel Trummer (Cornell)  
Ioana Manolescu (INRIA and Institut Polytechnique de Paris)  
Ippokratis Pandis (Amazon)  
Ishtiyaque Ahmad (University of California, Santa Barbara)  
Jae-Gil Lee (KAIST)  
Jana Giceva (TU Munich)  
Jeffrey Xu Yu (Chinese University of Hong Kong)  
Jens Teubner (TU Dortmund University)  
Jia Zou (Arizona State University)  
Jian Pei (Simon Fraser University)  
Jianguo Wang (Purdue University)  
Jiannan Wang (Simon Fraser University)  
Jianxin Li (Deakin University)  
Jianye Yang (Central South University)  
Jiwon Seo (Hanyang University)  
Johannes Gehrke (Microsoft)  
Jorge Arnulfo Quiane Ruiz (TU Berlin)  
Joseph Near (University of Vermont)  
Junhu Wang (Griffith University)  
Kaiping Zheng (National University of Singapore)  
Kangfei Zhao (The Chinese University of Hong Kong)  
Karima Echihabi (Mohammed VI Polytechnic University)  
Katja Hose (Aalborg University)  
Kenneth A Ross (Columbia University)  
Kostas Zoumpatianos (Snowflake Computing)  
Lei Zou (Peking University)  
Leopoldo Bertossi (Universidad Adolfo Ibanez)  
Li Xiong (Emory University)  
Lianke Qin (University of California, Santa Barbara)  
Lijun Chang (The University of Sydney)  
Lin Ma (Carnegie Mellon University)  
Long Yuan (Nanjing University of Science and Technology)  
Lu Qin (UTS)  
Luciano Barbosa (Universidade Federal de Pernambuco)  
Marcelo Arenas (Universidad Católica & IMFD)  
Maria Luisa Sapino (U. Torino)  
Matteo Lissandrini (Aalborg University)  
Matthias Boehm (Graz University of Technology)  
Matthias Renz (University of Kiel)  
Max Heimerl (Snowflake)  
Maximilian Schleich (University of Washington)  
Meihui Zhang (Beijing Institute of Technology)  
Melanie Herschel (Universität Stuttgart)  
Michael Abebe (University of Waterloo)  
Min Xie (Instacart)  
Mirella M Moro (Universidade Federal de Minas Gerais)  
Mohamed Sarwat (Arizona State University)  
Mohammad Dashti (MongoDB)  
Mohammad Javad Amiri (University of Pennsylvania)  
Mohammad Sadoghi (University of California, Davis)  
Muhammad Aamir Cheema (Monash University)

Nikita Bhutani (Megagon Labs)  
 Oliver A Kennedy (University at Buffalo, SUNY)  
 Panos K. Chrysanthis (University of Pittsburgh)  
 Paolo Missier (Newcastle University)  
 Parth Nagarkar (NMSU)  
 Paul Groth (University of Amsterdam)  
 Peng CHENG (East China Normal University)  
 Peter Pietzuch (Imperial College London)  
 Pierangela Samarati (Universita delgi Studi di Milano)  
 Pinar Karagoz (METU, Turkey)  
 Pinar Tozun (IT University of Copenhagen)  
 Prithu Banerjee (UBC)  
 Raoni Lourenço (New York University)  
 Raul Castro Fernandez (UChicago)  
 Ravi Ramamurthy (Microsoft)  
 Raymond Chi-Wing Wong (Hong Kong University of Science and Technology)  
 Renata Borovica-Gajic (University of Melbourne)  
 Reynold Cheng (The University of Hong Kong)  
 Rui Mao (Shenzhen University)  
 Ruoming Jin (Kent State University)  
 Sai Wu (Zhejiang University)  
 Sainyam Galhotra (University of Chicago)  
 Sanjay Krishnan (University of Chicago)  
 Sanjib Kumar Das (Google)  
 Sayan Ranu (IIT Delhi)  
 Sebastian Link (University of Auckland)  
 Semih Salihoglu (University of Waterloo)  
 Senjuti Basu Roy (New Jersey Institute of Technology)  
 Sergey Melnik (Google)  
 Shantanu Sharma (New Jersey Institute of Technology)  
 Shaoxu Song (Tsinghua University)  
 Sheng Wang (New York University)  
 Shimin Chen (Chinese Academy of Sciences)  
 Shumo Chu (University of California, Santa Barbara)  
 Shweta Jain (University of Illinois, Urbana-Champaign)  
 Sibow Wang (The Chinese University of Hong Kong)  
 Srinivasan Keshav (University of Cambridge)  
 Steffen Zeuch (DFKI GmbH)  
 Steven E Whang (KAIST)  
 Subarna Chatterjee (Harvard University)  
 Sudip Roy (Google)  
 Supun C Nakandala (University of California, San Diego)  
 Tamer Özsu (University of Waterloo)  
 Tarique A Siddiqui (Microsoft Research)  
 Thomas Heinis (Imperial College)  
 Thomas Neumann (TUM)  
 Tianzheng Wang (Simon Fraser University)  
 Tien Tuan Anh Dinh (Singapore University of Technology and Design)  
 Tilmann Rabl (HPI, University of Potsdam)  
 Ting Yu (Qatar Computing Research Institute)  
 Torben Bach Pedersen (Aalborg University)  
 Torsten Grust (Universität Tübingen)  
 Umar Farooq Minhas (Microsoft Research)  
 Vasiliki Kalavri (Boston University)  
 Verena Kantere (National Technical University of Athens)  
 Victor Zakhary (Oracle)  
 Vivek Narasayya (Microsoft Research)  
 Vraj Shah (University of California, San Diego)  
 Walid G Aref (Purdue)  
 Wasay Abdul (Harvard)  
 Wei Wang (Hong Kong University of Science and Technology (Guangzhou))  
 Wei Lu (Renmin university of china)  
 Weiren Yu (University of Warwick)  
 Wen Hua (The University of Queensland)  
 Wolfgang Lehner (TU Dresden)  
 Xi He (University of Waterloo)  
 Xiang Lian (Kent State University)  
 Xiao Qin (IBM Research)  
 Xiaofei Zhang (University of Memphis)  
 Xiaokui Xiao (National University of Singapore)  
 Xiaolan Wang (Megagon Labs)  
 Xiaoyang Wang (Zhejiang Gongshang University)  
 Xin Huang (Hong Kong Baptist University)  
 Yael Amsterdamer (Bar-Ilan university)  
 Yanyan Shen (Shanghai Jiao Tong University)  
 Ye Yuan (Northeastern University)  
 Yeye He (Microsoft Research)  
 Yi Chen (NJIT)  
 Yi Lu (MIT)  
 Yikai Zhang (Chinese University of Hong Kong)  
 Yinan Li (Microsoft Research)  
 Ying Zhang (University of Technology Sydney)  
 Yongxin Tong (Beihang University)  
 Yuanyuan Zhu (Wuhan University)  
 Yue Wang (Shenzhen Institute of Computing Sciences, Shenzhen University)  
 Yufei Tao (Chinese University of Hong Kong)  
 Yuliang Li (Megagon Labs)  
 Yuncheng Wu (National University of Singapore)  
 Yunjun Gao (Zhejiang University)  
 Yuval Moskovitch (University of Michigan)  
 Zhifeng Bao (RMIT University)  
 Zhongle Xie (Zhejiang University)  
 Zi Huang (University of Queensland)  
 Ziawasch Abedjan (Leibniz Universität Hannover)  
 Zohar Karnin (Amazon)  
 Zsolt István (IT University of Copenhagen)

## **INDUSTRIAL TRACK CHAIRS AND REVIEWERS - Vol. 15**

### **Industrial Track PC Chairs**

C. Mohan (Tsinghua University)  
Eric Simon (SAP)

### **Industrial Track Reviewers**

Allison Holloway (Oracle, USA)  
Alon Halevy (Facebook, USA)  
Amol Deshpande (University of Maryland and WireWheel.io, USA)  
Ashwin Machanavajjhala (Tumult Labs and Duke University, USA)  
Hanuma Kodavalla (Microsoft, USA)  
Ihab Ilyas (Apple, USA & University of Waterloo, Canada)  
Jianjun Chen (ByteDance, USA)  
Juan Sequeda (data.world, USA)  
Karthik Ramachandra (Microsoft, India)  
Lanjun Wang (Tianjin University, China)  
Laura Haas (University of Massachusetts at Amherst, USA)  
Lyublena Antova (Meta, USA)  
Matei Zaharia (Stanford University & Databricks, USA)  
Matthias Boehm (Graz University of Technology, Austria)  
Michael Carey (UC Irvine, USA)  
Michaela Hardt (Amazon Web Services, USA)  
Mohamed Soliman (Apple, USA)  
Mourad Ouzzani (Qatar Computing Research Institute, HBKU, Qatar)  
Nesime Tatbul (Intel Labs and MIT, USA)  
Norman May (SAP, Germany)  
Oktie Hassanzadeh (IBM Research, USA)  
Per-Ake Larson (University of Waterloo, Canada)  
Prasanta Ghosh (Microsoft, USA)  
Ryan Johnson (Databricks, USA)  
Sonia Bergamaschi (University of Modena & Reggio Emilia and DataRiver, Italy)  
Stefan Mandl (Exasol, Germany)  
Sudipto Das (Amazon Web Services, USA)  
Tyler Akidau (Snowflake, USA)  
Wei Wang (ByteDance, Singapore)  
Yannis Katsis (IBM Research, USA)  
Yingjun Wu (Singularity Data, USA)

## DEMONSTRATION TRACK CHAIRS AND REVIEWERS - Vol. 15

### Demonstration Track PC Chairs

Eduard Dragut (Temple University)  
Immanuel Trummer (Cornell University)

### Demonstration Track Reviewers

Alexander van Renen (Friedrich-Alexander-Universität Erlangen-Nürnberg)  
Altigran Soares da Silva (Universidade Federal do Amazonas)  
Amir Gilad (Duke University)  
Amit Somech (Bar-Ilan University)  
Andrew Crotty (CMU)  
Avigdor Gal (Technion – Israel Institute of Technology)  
Behrooz Omidvar-Tehrani (Amazon)  
Bin Cui (PKU)  
Chenglong Wang (Microsoft Research)  
Cong Yan (Microsoft Research)  
Dan Olteanu (University of Zurich)  
El Kindi Rezig (MIT)  
Enzo Veltri (Università della Basilicata)  
Gabriel Ghinita (University of Massachusetts at Boston)  
Georgia Troullinou (FORTH-ICS)  
Harish Doraiswamy (Microsoft Research)  
Hiroaki Shiokawa (University of Tsukuba)  
Jarek Szlichta (Ontario Tech University)  
Jialin Ding (MIT)  
Jiangwei Zhang (Tencent)  
Jiaqi Yan (Snowflake)  
Ju Fan (Renmin University of China)  
Jyoti Leeka (Microsoft Research)  
Krishna Kantikiran Pasupuleti (Oracle)  
Laurel Orr (Stanford)  
Lei Li (The University of Queensland)  
Leonardo Andrade Ribeiro (Federal University of Goiás)  
Lihong He (Temple University)  
Luciano Barbosa (Universidade Federal de Pernambuco)  
Lukasz Golab (University of Waterloo)  
Maeda Hanafi (IBM Almaden Research Center)  
Manisha Luthra (Technical University of Darmstadt)  
Maristela Terto de Holanda (University of Brasilia)  
Mayuresh Prakash Kunjir (QCRI)  
Milos Nikolic (University of Edinburgh)  
Mourad Ouzzani (Qatar Computing Research Institute, HBKU)  
Nan Tang (QCRI)  
Nantia Makrynioti (CWI)  
Olga Poppe (Microsoft Research)  
Pedro Holanda (CWI)  
Raja Appuswamy (Eurecom)  
Ramon Lawrence (The University of British Columbia)  
Raymond Wong (Hong Kong University of Science and Technology)  
Rebecca Taft (Cockroach Labs)  
Renata Borovica-Gajic (University of Melbourne)  
Roe Shraga (Northeastern University)  
Rong Zhu (Alibaba Group)  
Sainyam Galhotra (University of Chicago)  
Sang-Wook Kim (Hanyang University)  
Saravanan Thirumuruganathan (QCRI)  
Shivaram Venkataraman (University of Wisconsin, Madison)  
Simon Razniewski (MPI)  
Subarna Chatterjee (Harvard)  
Subhadeep Sarkar (Boston University)  
Tiago Eugenio de Melo (Universidade do Estado do Amazonas)  
Tsz Nam Chan (Hong Kong Baptist University)  
Varun Pandey (TU Berlin)  
Venkatesh Emani (Microsoft Gray Systems Lab)  
Verena Kantere (National Technical University of Athens)  
Vladislav Shkapenyuk (AT&T Labs)  
Walter Cai (Snowflake)  
Weijie Zhao (Baidu Research)  
Y.C. Tay (National University of Singapore)  
Yaron Kanza (AT&T Labs)  
Yingxia Shao (Beijing University of Posts and Telecommunications)  
Yiru Chen (Columbia University)  
Yiwen Zhu (Microsoft)  
Yuto Hayamizu (The University of Tokyo)  
Yuval Moskovitch (University of Michigan)  
Zainab Abbas (KTH)  
Zeyuan Shang (Einblick)  
Zheguang Zhao (Technical University of Darmstadt)  
Zhi Yang (PKU)

## LETTER FROM THE EDITORS IN CHIEF

Welcome to the twelfth issue of Proceedings of the VLDB Endowment (PVLDB), Volume 15. PVLDB provides a high-quality 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 weeks, 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 submissions covering a diverse set of data management topics, including, cloud databases, data integration, machine learning and databases, distributed transactions, query processing and optimization, graph analytics, scalable data science, and distributed systems. The review process was carried out by 34 Associate Editors, who coordinated the reviews, and 172 expert researchers on the Review Board of Volume 15 who provided high quality reviews.

This issue contains 22 industrial papers, 46 demo papers, 9 tutorials, 2 panel abstracts, 2 keynote talk articles, and 3 endowment award articles.

The industry track co-chairs are C. Mohan and Eric Simon. The industry program committee consists of 31 members who are based in Austria, Canada, China, Germany, India, Italy, Qatar, Singapore and USA. The PC members refereed the 60 papers that were submitted, covering topics on cloud considerations, compression, distributed databases, encryption, graph databases, hardware acceleration, HTAP, non-volatile memory, spatial data, benchmarking, confidential computing, data lakes, distributed file systems, IoT, machine learning, physical database design, query optimization and processing.

The demonstration track is chaired by Immanuel Trummer and Eduard Dragut, and the program committee consisted of 72 researchers. 46 demonstrations are accepted out of a total of 106 submissions. The accepted contributions cover a variety of topics, including systems and interfaces for data science and data cleaning, novel approaches for query processing and benchmark generation, as well as specialized systems and interfaces for graph data, spatial data, time series, and visual data.

The tutorial track co-chairs are Yoshiharu Ishikawa and Laks V.S. Lakshmanan. They assembled a program committee consisting of 10 experts including 3 women, spanning 6 countries and covering a wide range of topics. Of the 18 proposals that were received, the program committee selected 9 tutorials, judged to be high quality and timely, for presentation at the conference. Topics featured by the selected tutorials are Machine Programming, Cloud Databases, Transformers for Tabular Data, Polyglot Data Management, Querying Graph Structured Relations, Indexing on Persistent Memory, Discovering Densest Subgraphs, Unified Data Analytics, and Harnessing Very Large Language Models for Data Management.

The panel track co-chairs are Azza Abouzied, Peter Bailis and AnHai Doan, who selected two panels. The first panel on "Cloud Data" is organized by Surajit Chaudhuri, and the second panel on "Startups by Database Researchers" is organized by C. Mohan. The abstracts for these panels are included in this issue.

VLDB 2022 will have 3 keynote talks. This issue contains the extended abstract for two of these three talks: Wenfei Fan talks about "Big Graphs: Challenges and Opportunities", and Sihem Amir-Yahia talks about "Towards AI-Powered Data-Driven Education".

VLDB Early Career Research Contribution Award this year is awarded to Sudeepa Roy and the VLDB Women in Database Research Award is awarded to Fatma Özcan. Finally, the 2022 VLDB Test of Time Award is awarded to Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, Tianyi Wu for their VLDB 2011 article entitled "PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks". This issue also contains extended abstracts for these three awards.

All these papers will be presented at the 48th International Conference on Very Large Data Bases (VLDB 2022) between September 5th and September 9th, 2022. The conference will be run in a hybrid mode, including participants attending in person in Sydney, Australia, as well as attending remotely. We sincerely thank all the authors for submitting their work and all the reviewers and the track chairs for their outstanding service in reviewing

the submissions. We hope that the readers will find the selected papers engaging, and thought provoking, providing valuable insights and inspiring follow-up research.

All papers will be presented at the 2022 Conference on Very Large Databases (VLDB 2022) in Sydney, Australia. We hope you enjoy reading this issue and look forward to seeing you in Sydney!

Fatma Özcan, Juliana Freire and Xuemin Lin  
Editors-in-Chief of PVLDB Volume 15  
Program Chairs for VLDB 2022