

# **49th International Conference on Very Large Data Bases (VLDB 2023)**

Proceedings of the VLDB Endowment Volume 16

Vancouver, Canada  
28 August – 1 September 2023

Part 1 of 6

**Editors:**

**Georgia Koutrika  
Jun Yang**

ISBN: 978-1-7138-8478-1

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2022) by 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-ncnd/4.0/>. For any use beyond those covered by this license, obtain permission by emailing [info@vldb.org](mailto:info@vldb.org).

Printed with permission by Curran Associates, Inc. (2024)

For permission requests, please contact VLDB Endowment at the address below.

VLDB Endowment  
Attn: Marianne Siroker  
353 Serra Mall – Stanford University  
Stanford CA 94305

Phone: (650) 723-0872

Fax: (650) 725-2588

[info@vldb.org](mailto:info@vldb.org).

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)



# Proceedings of the VLDB Endowment

Volume 16 – September 2022

**Editors in Chief:**

Georgia Koutrika and Jun Yang

**Associate Editors:**

Alkis Simitsis, Amol Deshpande, Angela Bonifati, Ashwin Machanavajjhala, Badrish Chandramouli,  
Boris Glavic, Ce Zhang, Cyrus Shahabi, Dan Olteanu, Eric Lo, Evaggelia Pitoura, Evimaria Terzi,  
Gustavo Alonso, Helen (Zi) Huang, Hong Cheng, Kenneth Ross, Khuzaima Daudjee, Kyuseok Shim,  
Letizia Tanca, Lucian Popa, Magdalena Balazinska, Meihui Zhang, Neoklis Polyzotis,  
Nesime Tatbul, Nikos Mamoulis, Rachel Pottinger, Wenjie Zhang, Wolfgang Gatterbauer,  
Wook-Shin Han, Xiaokui Xiao, Yannis Velegrakis, Yanyan Shen, Yi Chen, Yongxin Tong, Zhifeng Bao

**Publication Editors:**

Manos Athanassoulis, Kostas Stefanidis, Ju Fan

PVLDB – Proceedings of the VLDB Endowment

All papers published in this issue will be presented at the 49th International Conference on Very Large Data Bases, Vancouver, Canada, 2023.

## **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).

## TABLE OF CONTENTS

### Research Papers

C5: Cloned Concurrency Control That Always Keeps Up .....	1
<i>Jeffrey Helt, Abhinav Sharma, Daniel J Abadi, Wyatt Lloyd, Jose Faleiro</i>	
The Case for Distributed Shared-Memory Databases with RDMA-Enabled Memory Disaggregation .....	15
<i>Ruihong Wang, Jianguo Wang, Stratos Idreos, Tamer Özsu, Walid G Aref</i>	
FlexChain: An Elastic Disaggregated Blockchain.....	23
<i>Chenyuan Wu, Mohammad Javad Amiri, Jared Asch, Heena Nagda, Qizhen Zhang, Boon Thau Loo</i>	
MiCS: Near-linear Scaling for Training Gigantic Model on Public Cloud.....	37
<i>Zhen Zhang, Shuai Zheng, Yida Wang, Justin Chiu, George Karypis, Trishul A Chilimbi, Mu Li, Xin Jin</i>	
Privacy-preserving Cooperative Online Matching over Spatial Crowdsourcing Platforms .....	51
<i>Yi Yang, Yurong Cheng, Ye Yuan, Guoren Wang, Lei Chen, Yongjiao Sun</i>	
Coresets over Multiple Tables for Feature-rich and Data-efficient Machine Learning.....	64
<i>Jiayi Wang, Chengliang Chai, Nan Tang, Jiabin Liu, Guoliang Li</i>	
STARRY: Multi-master Transaction Processing on Semi-leader Architecture.....	77
<i>Zihao Zhang, Huiqi Hu, Xuan Zhou, Jiang Wang</i>	
SIFTER: Space-Efficient Value Iteration for Finite-Horizon MDPs .....	90
<i>Konstantinos Skitsas, Ioannis G Papageorgiou, Mohammad Sadegh Talebi, Vasiliki Kantere, Michael Katehakis, Panagiotis Karras</i>	
TreeLine: An Update-In-Place Key-Value Store for Modern Storage .....	99
<i>Geoffrey Yu, Markos Markakis, Andreas Kipf, Per-ake Larson, Umar Farooq Minhas, Tim Kraska</i>	
DPXPlain: Privately Explaining Aggregate Query Answers .....	113
<i>Yuchao Tao, Amir Gilad, Ashwin Machanavajjhala, Sudeepa Roy</i>	

## TABLE OF CONTENTS

### Research Papers

Efficient Maximum k-Plex Computation over Large Sparse Graphs.....	127
<i>Lijun Chang, Mouyi Xu, Darren Strash</i>	
Online Schema Evolution is (almost) Free for Snapshot Databases.....	140
<i>Tianxun Hu, Tianzheng Wang, Qingqing Zhou</i>	
LIDER: An Efficient High-dimensional Learned Index for Large-scale Dense Passage Retrieval .....	154
<i>Yifan Wang, Haodi Ma, Daisy Zhe Wang</i>	
Models and Mechanisms for Spatial Data Fairness.....	167
<i>Sina Shaham, Gabriel Ghinita, Cyrus Shahabi</i>	
Influence Maximization in Real-World Closed Social Networks.....	180
<i>Shixun Huang, Wenqing Lin, Zhifeng Bao, Jiachen Sun</i>	
Time2Feat: Learning Interpretable Representations for Multivariate Time Series Clustering.....	193
<i>Angela Bonifati, Francesco Del Buono, Francesco Guerra, Donato Tiano</i>	
OpBoost: A Vertical Federated Tree Boosting Framework Based on Order-Preserving Desensitization .....	202
<i>Xiao Chen Li, Yuke Hu, Weiran Liu, Hanwen Feng, Li Peng, Yuan Hong, Kui Ren, Zhan Qin</i>	
HMAB: Self-Driving Hierarchy of Bandits for Integrated Physical Database Design Tuning .....	216
<i>R. Malinga Perera, Bastian Oetomo, Benjamin I. P. Rubinstein, Renata Borovica-Gajic</i>	
Erebus: Explaining the Outputs of Data Streaming Queries .....	230
<i>Dimitris Palyvos-Giannas, Katerina Tzompanaki, Marina Papatriantafilou, Vincenzo Gulisano</i>	
PLIN: A Persistent Learned Index for Non-Volatile Memory with High Performance and Instant Recovery ...	243
<i>Zhou Zhang, Zhaole Chu, Peiquan Jin, Yongping Luo, Xike Xie, Shouhong Wan, Yun Luo, Xufei Wu, Peng Zou, Chunyang Zheng, Guoan Wu, Andy Rudoff</i>	
Fries: Fast and Consistent Runtime Reconfiguration in Dataflow Systems with Transactional Guarantees.....	256
<i>Zuozhi Wang, Shengquan Ni, Avinash Kumar, Chen Li</i>	
Fast Approximate Denial Constraint Discovery.....	269
<i>Renjie Xiao, Zijing Tan, Haojin Wang, Shuai Ma</i>	
Frequency Domain Data Encoding in Apache IoTDB.....	282
<i>Haoyu Wang, Shaoxu Song</i>	
Happiness Maximizing Sets under Group Fairness Constraints.....	291
<i>Jiping Zheng, Yuan Ma, Wei Ma, Yanhao Wang, Xiaoyang Wang</i>	
SHiFT: An Efficient, Flexible Search Engine for Transfer Learning .....	304
<i>Cedric Renggli, Xiaozhe Yao, Luka Kolar, Luka Rimanic, Ana Klimovic, Ce Zhang</i>	

Satisfying Complex Top-k Fairness Constraints by Preference Substitutions .....	317
<i>Md. Mouinul Islam, Dong Wei, Baruch Schieber, Senjuti Basu Roy</i>	
SyncSignature: A Simple, Efficient, Parallelizable Framework for Tree Similarity Joins .....	330
<i>Nikolai Karpov, Qin Zhang</i>	
Approximating Probabilistic Group Steiner Trees in Graphs .....	343
<i>Shuang Yang, Yahui Sun, Jiesong Liu, Xiaokui Xiao, Rong-Hua Li, Zhewei Wei</i>	
Space-Efficient Random Walks on Streaming Graphs .....	356
<i>Serafeim Papadias, Zoi Kaoudi, Jorge-Arnulfo Quiané-Ruiz, Volker Markl</i>	
PromptEM: Prompt-tuning for Low-resource Generalized Entity Matching .....	369
<i>Pengfei Wang, Xiaocan Zeng, Lu Chen, Fan Ye, Yuren Mao, Junhao Zhu, Yunjun Gao</i>	
Cornus: Atomic Commit for a Cloud DBMS with Storage Disaggregation .....	379
<i>Zhihan Guo, Xinyu Zeng, Kan Wu, Wuh-Chwen Hwang, Ziwei Ren, Xiangyao Yu, Mahesh Balakrishnan, Philip A. Bernstein</i>	

## TABLE OF CONTENTS

### Research Papers

Route Travel Time Estimation on A Road Network Revisited: Heterogeneity, Proximity, Periodicity and Dynamicity .....	393
<i>Haitao Yuan, Guoliang Li, Zhifeng Bao</i>	
Serving and Optimizing Machine Learning Workflows on Heterogeneous Infrastructures .....	406
<i>Yongji Wu, Matthew Lentz, Danyang Zhuo, Yao Lu</i>	
Computing Rule-Based Explanations by Leveraging Counterfactuals.....	420
<i>Zixuan Geng, Maximilian Schleich, Dan Suciu</i>	
Self-supervised and Interpretable Data Cleaning with Sequence Generative Adversarial Networks .....	433
<i>Jinfeng Peng, Derong Shen, Nan Tang, Tieying Liu, Yue Kou, Tiezheng Nie, Hang Cui, Ge Yu</i>	
Optimizing Video Analytics with Declarative Model Relationships.....	447
<i>Francisco Romero, Johann Hauswald, Aditi Partap, Daniel Kang, Matei Zaharia, Christos Kozyrakis</i>	
Spade: A Real-Time Fraud Detection Framework on Evolving Graphs.....	461
<i>Jiaxin Jiang, Yuan Li, Bingsheng He, Bryan Hooi, Jia Chen, Johan Kok Zhi Kang</i>	
Galvatron: Efficient Transformer Training over Multiple GPUs Using Automatic Parallelism.....	470
<i>Xupeng Miao, Yujie Wang, Youhe Jiang, Chunan Shi, Xiaonan Nie, Hailin Zhang, Bin Cui</i>	
iDeaL: A Deep Learning Framework for Detecting Highly Imbalanced Interictal Epileptiform Discharges ...	480
<i>Qitong Wang, Stephen Whitmarsh, Vincent Navarro, Themis Palpanas</i>	
Dalton: Learned Partitioning for Distributed Data Streams .....	491
<i>Eleni Zapridou, Ioannis Mytilinis, Anastasia Ailamaki</i>	
FirmTruss Community Search in Multilayer Networks .....	505
<i>Ali Behrouz, Farnoosh Hashemi, Laks V.S. Lakshmanan</i>	
Efficient Triangle-Connected Truss Community Search In Dynamic Graphs .....	519
<i>Tianyang Xu, Zhao Lu, Yuanyuan Zhu</i>	
Can Learned Models Replace Hash Functions? .....	532
<i>Ibrahim Sabek, Kapil Vaidya, Dominik Horn, Andreas Kipf, Michael Mitzenmacher, Tim Kraska</i>	
TOD: GPU-accelerated Outlier Detection via Tensor Operations.....	546
<i>Yue Zhao, George H. Chen, Zhihao Jia</i>	
FILM: a Fully Learned Index for Larger-than-Memory Databases.....	561
<i>Chaohong Ma, Xiaohui Yu, Yifan Li, Xiaofeng Meng, Aishan Maolinyazi</i>	



## TABLE OF CONTENTS

### Research Papers

Cache Me If You Can: Accuracy-Aware Inference Engine for Differentially Private Data Exploration.....	574
<i>Miti Mazmudar, Thomas Humphries, Jiaxiang Liu, Matthew Rafuse, Xi He</i>	
Range Search over Encrypted Multi-Attribute Data.....	587
<i>Francesca Falzon, Evangelia Anna Markatou, Zachary Espiritu, Roberto Tamassia</i>	
HEDA: Multi-Attribute Unbounded Aggregation over Homomorphically Encrypted Database .....	601
<i>Xuanle Ren, Le Su, Zhen Gu, Sheng Wang, Feifei Li, Yuan Xie, Song Bian, Chao Li, Fan Zhang</i>	
Density Personalized Group Query .....	615
<i>Chih-Ya Shen, Shao-Heng Ko, Guang-Siang Lee, Wang-Chien Lee, De-Nian Yang</i>	
Nezha: Deployable and High-Performance Consensus Using Synchronized Clocks.....	629
<i>Jinkun Geng, Anirudh Sivaraman, Balaji Prabhakar, Mendel Rosenblum</i>	
Pantheon: Private Retrieval from Public Key-Value Store .....	643
<i>Ishtiyaque Ahmad, Divyakant Agrawal, Amr El Abbadi, Trinabh Gupta</i>	
Bayesian Sketches for Volume Estimation in Data Streams .....	657
<i>Francesco Da Dalt, Simon Scherrer, Adrian Perrig</i>	
Waffle: A Workload-Aware and Query-Sensitive Framework for Disk-Based Spatial Indexing.....	670
<i>Moin Hussain Moti, Panagiotis Simatis, Dimitris Papadias</i>	
Fast Algorithms for Denial Constraint Discovery.....	684
<i>Eduardo H. M. Pena, Fabio Porto, Felix Naumann</i>	
Toward Quantity-of-Interest Preserving Lossy Compression for Scientific Data.....	697
<i>Pu Jiao, Sheng Di, Hanqi Guo, Kai Zhao, Jiannan Tian, Dingwen Tao, Xin Liang, Franck Cappelto</i>	
Scalable Graph Convolutional Network Training on Distributed-Memory Systems .....	711
<i>Gunduz Vehbi Demirci, Aparajita Haldar, Hakan Ferhatosmanoglu</i>	
Motiflets - Simple and Accurate Detection of Motifs in Time Series .....	725
<i>Patrick Schäfer, Ulf Leser</i>	
Can Foundation Models Wrangle Your Data? .....	738
<i>Avanika Narayan, Ines Chami, Laurel Orr, Christopher Ré</i>	
M2Bench: A Database Benchmark for Multi-Model Analytic Workloads .....	747
<i>Bogyong Kim, Kyoseung Koo, Undraa Enkhbat, Sohyun Kim, Juhun Kim, Bongki Moon</i>	
Parallelism-Optimizing Data Placement for Faster Data-Parallel Computations .....	760
<i>Nirvik Baruah, Peter Kraft, Fiodar Kazhamiaka, Peter Bailis, Matei Zaharia</i>	
SubStrat: A Subset-Based Optimization Strategy for Faster AutoML.....	772
<i>Teddy Lazebnik, Amit Somech, Abraham Itzhak Weinberg</i>	

MultiBiSage: A Web-Scale Recommendation System Using Multiple Bipartite Graphs at Pinterest .....	781
<i>Saket Gurukar, Nikil Pancha, Andrew Zhai, Eric Kim, Samson Hu, Srinivasan Parthasarathy, Charles Rosenberg, Jure Leskovec</i>	
TokenJoin: Efficient Filtering for Set Similarity Join with Maximum Weighted Bipartite Matching.....	790
<i>Alexandros Zeakis, Dimitrios Skoutas, Dimitris Sacharidis, Odysseas Papapetrou, Manolis Koubarakis</i>	
Quasi-stable Coloring for Graph Compression: Approximating Max-Flow, Linear Programs, and Centrality .	803
<i>Moe Kayali, Dan Suciu</i>	
Multi-Analyst Differential Privacy for Online Query Answering.....	816
<i>David Pujol, Albert Sun, Brandon Fain, Ashwin Machanavajhala</i>	
Excalibur: A Virtual Machine for Adaptive Fine-grained JIT-Compiled Query Execution based on VOILA .....	829
<i>Tim Gubner, Peter Boncz</i>	
Differentially Oblivious Relational Database Operators.....	842
<i>Lianke Qin, Rajesh Jayaram, Elaine Shi, Zhao Song, Danyang Zhuo, Shumo Chu</i>	
Keep CALM and CRDT On.....	856
<i>Shadaj Laddad, Conor Power, Mae Milano, Alvin Cheung, Natacha Crooks, Joseph M. Hellerstein</i>	
MQH: Locality Sensitive Hashing on Multi-level Quantization Errors for Point-to-Hyperplane Distances .....	864
<i>Kejing Lu, Yoshiharu Ishikawa, Chuan Xiao</i>	
The LDBC Social Network Benchmark: Business Intelligence Workload.....	877
<i>Gábor Szárnyas, Jack Waudby, Benjamin A. Steer, Dávid Szakállas, Altan Birler, Mingxi Wu, Yuchen Zhang, Peter Boncz</i>	
Making Cache Monotonic and Consistent.....	891
<i>Shuai An, Yang Cao</i>	
SkinnerMT: Parallelizing for Efficiency and Robustness in Adaptive Query Processing on Multicore Platforms .....	905
<i>Ziyun Wei, Immanuel Trummer</i>	
On Efficient Approximate Queries over Machine Learning Models.....	918
<i>Dujian Ding, Sihem Amer-Yahia, Laks Lakshmanan</i>	
Integrating Data Lake Tables .....	932
<i>Aamod Khatiwada, Roei Shraga, Wolfgang Gatterbauer, Renée J. Miller</i>	
PIM-tree: A Skew-resistant Index for Processing-in-Memory .....	946
<i>Hongbo Kang, Yiwei Zhao, Guy E. Blelloch, Laxman Dhulipala, Yan Gu, Charles Mcguffey, Phillip B. Gibbons</i>	
Web Record Extraction with Invariants .....	959
<i>Zhijia Chen, Weiyi Meng, Eduard Dragut</i>	
A Deep Generative Model for Trajectory Modeling and Utilization .....	973
<i>Yong Wang, Guoliang Li, Kaiyu Li, Haitao Yuan</i>	
L2chain: Towards High-performance, Confidential and Secure Layer-2 Blockchain Solution for Decentralized Applications.....	986
<i>Zihuan Xu, Lei Chen</i>	

## TABLE OF CONTENTS

### Research Papers

Auto-Tuning with Reinforcement Learning for Permissioned Blockchain Systems .....	1000
<i>Mingxuan Li, Yazhe Wang, Shuai Ma, Chao Liu, Dongdong Huo, Yu Wang, Zhen Xu</i>	
PetPS: Supporting Huge Embedding Models with Persistent Memory.....	1013
<i>Minhui Xie, Youyou Lu, Qing Wang, Yangyang Feng, Jiaqiang Liu, Kai Ren, Jiwu Shu</i>	
Extraction of Validating Shapes from very large Knowledge Graphs.....	1023
<i>Kashif Rabbani, Matteo Lissandrini, Katja Hose</i>	
Async-fork: Mitigating Query Latency Spikes Incurred by the Fork-based Snapshot Mechanism from the OS Level.....	1033
<i>Pu Pang, Gang Deng, Kaihao Bai, Quan Chen, Shixuan Sun, Bo Liu, Yu Xu, Hongbo Yao, Zhengheng Wang, Xiyu Wang, Zheng Liu, Zhuo Song, Yong Yang, Tao Ma, Minyi Guo</i>	
Change Propagation Without Joins.....	1046
<i>Qichen Wang, Xiao Hu, Binyang Dai, Ke Yi</i>	
FederatedScope: A Flexible Federated Learning Platform for Heterogeneity.....	1059
<i>Yuexiang Xie, Zhen Wang, Dawei Gao, Daoyuan Chen, Liuyi Yao, Weirui Kuang, Yaliang Li, Bolin Ding, Jingren Zhou</i>	
ACTA: Autonomy and Coordination Task Assignment in Spatial Crowdsourcing Platforms .....	1073
<i>Boyang Li, Yurong Cheng, Ye Yuan, Yi Yang, Qianqian Jin, Guoren Wang</i>	
FastFlow: Accelerating Deep Learning Model Training with Smart Offloading of Input Data Pipeline .....	1086
<i>Taegeon Um, Byungsoo Oh, Byeongchan Seo, Minhyeok Kweun, Goeun Kim, Woo-Yeon Lee</i>	
FARGO: Fast Maximum Inner Product Search via Global Multi-Probing .....	1100
<i>Xi Zhao, Bolong Zheng, Xiaomeng Yi, Xiaofan Luan, Charles Xie, Xiaofang Zhou, Christian S. Jensen</i>	
Optimistic Data Parallelism for FPGA-Accelerated Sketching.....	1113
<i>Martin Kiefer, Ilias Poulakis, Eleni Tzirita Zacharatou, Volker Markl</i>	
On the Risks of Collecting Multidimensional Data Under Local Differential Privacy .....	1126
<i>Héber H. Arcolezi, Sébastien Gams, Jean-François Couchot, Catuscia Palamidessi</i>	
Odyssey: A Journey in the Land of Distributed Data Series Similarity Search.....	1140
<i>Manos Chatzakis, Panagiota Fatourou, Eleftherios Kosmas, Themis Palpanas, Botao Peng</i>	
Anonymous Edge Representation for Inductive Anomaly Detection in Dynamic Bipartite Graphs.....	1154
<i>Lanting Fang, Kaiyu Feng, Jie Gui, Shanshan Feng, Aiqun Hu</i>	
Scalable Time-Range k-Core Query on Temporal Graphs.....	1168
<i>Junyong Yang, Ming Zhong, Yuanyuan Zhu, Tiejun Qian, Mengchi Liu, Jeffrey Xu Yu</i>	
High-Performance Row Pattern Recognition Using Joins.....	1181
<i>Erkang Zhu, Silu Huang, Surajit Chaudhuri</i>	

A Hierarchical Grouping Algorithm for the Multi-Vehicle Dial-a-Ride Problem ..... 1195  
*Kelin Luo, Alexandre M. Florio, Syamantak Das, Xiangyu Guo*

## TABLE OF CONTENTS

### Research Papers

Leveraging Application Data Constraints to Optimize Database-Backed Web Applications .....	1208
<i>Xiaoxuan Liu, Shuxian Wang, Mengzhu Sun, Sicheng Pan, Ge Li, Siddharth Jha, Cong Yan, Junwen Yang, Shan Lu, Alvin Cheung</i>	
Bringing Compiling Databases to RISC Architectures.....	1222
<i>Ferdinand Gruber, Maximilian Bandle, Alexis Engelke, Thomas Neumann, Jana Giceva</i>	
Blink-hash: An Adaptive Hybrid Index for In-Memory Time-Series Databases .....	1235
<i>Hokeun Cha, Xiangpeng Hao, Tianzheng Wang, Huanchen Zhang, Aditya Akella, Xiangyao Yu</i>	
A Design Space Exploration and Evaluation for Main-Memory Hash Joins in Storage Class Memory.....	1249
<i>Wentao Huang, Yunhong Ji, Xuan Zhou, Bingsheng He, Kian-Lee Tan</i>	
Efficient Black-box Checking of Snapshot Isolation in Databases.....	1264
<i>Kaile Huang, Si Liu, Zheng Chen, Hengfeng Wei, David Basin, Haixiang Li, Anqun Pan</i>	
Differentially Private Vertical Federated Clustering .....	1277
<i>Zitao Li, Tianhao Wang, Ninghui Li</i>	
Panagos: Chasing the Tails for Multidimensional Data Streams.....	1291
<i>Fuheng Zhao, Punnal Ismail Khan, Divyakant Agrawal, Amr El Abbadi, Arpit Gupta, Zaoxing Liu</i>	
VersaMatch: Ontology Matching with Weak Supervision.....	1305
<i>Jonathan Fürst, Mauricio Fadel Argerich, Bin Cheng</i>	
RECA: Related Tables Enhanced Column Semantic Type Annotation Framework.....	1319
<i>Yushi Sun, Hao Xin, Lei Chen</i>	
Zebra: When Temporal Graph Neural Networks Meet Temporal Personalized PageRank .....	1332
<i>Yiming Li, Yanyan Shen, Lei Chen, Mingxuan Yuan</i>	
Efficient Approximation of Certain and Possible Answers for Ranking and Window Queries over Uncertain Data .....	1346
<i>Su Feng, Boris Glavic, Oliver Kennedy</i>	
GlassDB: An Efficient Verifiable Ledger Database System Through Transparency.....	1359
<i>Cong Yue, Tien Tuan Anh Dinh, Zhongle Xie, Meihui Zhang, Gang Chen, Beng Chin Ooi, Xiaokui Xiao</i>	
Efficient Distributed Transaction Processing in Heterogeneous Networks.....	1372
<i>Qian Zhang, Jingyao Li, Hongyao Zhao, Quanqing Xu, Wei Lu, Jinliang Xiao, Fusheng Han, Chuanhui Yang, Xiaoyong Du</i>	
Auxo: A Scalable and Efficient Graph Stream Summarization Structure.....	1386
<i>Zhiguo Jiang, Hanhua Chen, Hai Jin</i>	

OneShotSTL: One-Shot Seasonal-Trend Decomposition For Online Time Series Anomaly Detection And Forecasting.....	1399
<i>Xiao He, Ye Li, Jian Tan, Bin Wu, Feifei Li</i>	
Cloud Analytics Benchmark.....	1413
<i>Alexander Van Renen, Viktor Leis</i>	
Scalable and Robust Snapshot Isolation for High-Performance Storage Engines .....	1426
<i>Adnan Alhomssi, Viktor Leis</i>	
FLARE: A Fast, Secure, and Memory-Efficient Distributed Analytics Framework (Flavor: Systems) .....	1439
<i>Xiang Li, Fabing Li, Mingyu Gao</i>	
NV-SQL: Boosting OLTP Performance with Non-Volatile DIMMs .....	1453
<i>Mijin An, Jonghyeok Park, Tianzheng Wang, Beomseok Nam, Sang-Won Lee</i>	
Lero: A Learning-to-Rank Query Optimizer.....	1466
<i>Rong Zhu, Wei Chen, Bolin Ding, Xingguang Chen, Andreas Pfadler, Ziniu Wu, Jingren Zhou</i>	
Deploying Computational Storage for HTAP DBMSs Takes More Than Just Computation Offloading.....	1480
<i>Kitaeek Lee, Insoon Jo, Jaechan Ahn, Hyuk Lee, Hwang Lee, Woong Sul, Hyungsoo Jung</i>	
Transactional Panorama: A Conceptual Framework for User Perception in Analytical Visual Interfaces .....	1494
<i>Dixin Tang, Alan Fekete, Indranil Gupta, Aditya G. Parameswaran</i>	
Sparkly: A Simple yet Surprisingly Strong TF/IDF Blocker for Entity Matching.....	1507
<i>Derek Paulsen, Yash Govind, Anhai Doan</i>	
Robust Query Driven Cardinality Estimation under Changing Workloads .....	1520
<i>Parimarjan Negi, Ziniu Wu, Andreas Kipf, Nesime Tatbul, Ryan Marcus, Sam Madden, Tim Kraska, Mohammad Alizadeh</i>	
CatSQL: Towards Real World Natural Language to SQL Applications .....	1534
<i>Han Fu, Chang Liu, Bin Wu, Feifei Li, Jian Tan, Jianling Sun</i>	
Elpis: Graph-Based Similarity Search for Scalable Data Science .....	1548
<i>Ilias Azizi, Karima Echihabi, Themis Palpanas</i>	
Dotori: A Key-Value SSD Based KV Store .....	1560
<i>Carl Duffy, Jaehoon Shim, Sang-Hoon Kim, Jin-Soo Kim</i>	
PreFair: Privately Generating Justifiably Fair Synthetic Data.....	1573
<i>David Pujol, Amir Gilad, Ashwin Machanavajjhala</i>	
Explaining Dataset Changes for Semantic Data Versioning with Explain-Da-V .....	1587
<i>Roe Shraga, Renée J. Miller</i>	

## TABLE OF CONTENTS

### Research Papers

DBSP: Automatic Incremental View Maintenance for Rich Query Languages.....	1601
<i>Mihai Budiu, Tej Chajed, Frank Mcsherry, Leonid Ryzhyk, Val Tannen</i>	
SPG: Structure-Private Graph Database via SqueezePIR.....	1615
<i>Ling Liang, Jilan Lin, Zheng Qu, Ishtiyaque Ahmad, Fengbin Tu, Trinabh Gupta, Yufei Ding, Yuan Xie</i>	
InfiniStore: Elastic Serverless Cloud Storage.....	1629
<i>Jingyuan Zhang, Ao Wang, Xiaolong Ma, Benjamin Carver, Nicholas John Newman, Ali Anwar, Lukas Rupprecht, Vasily Tarasov, Dimitrios Skourtis, Feng Yan, Yue Cheng</i>	
Distributed Graph Embedding with Information-Oriented Random Walks.....	1643
<i>Peng Fang, Arijit Khan, Siqiang Luo, Fang Wang, Dan Feng, Zhenli Li, Wei Yin, Yuchao Cao</i>	
Secure Shapley Value for Cross-Silo Federated Learning.....	1657
<i>Shuyuan Zheng, Yang Cao, Masatoshi Yoshikawa</i>	
SODA: A Set of Fast Oblivious Algorithms in Distributed Secure Data Analytics .....	1671
<i>Xiang Li, Nuozhou Sun, Yunqian Luo, Mingyu Gao</i>	
GriDB: Scaling Blockchain Database via Sharding and Off-Chain Cross-Shard Mechanism .....	1685
<i>Zicong Hong, Song Guo, Enyuan Zhou, Wuhui Chen, Huawei Huang, Albert Zomaya</i>	
SUFF: Accelerating Subgraph Matching with Historical Data .....	1699
<i>Xun Jian, Zhiyuan Li, Lei Chen</i>	
When Database Meets New Storage Devices: Understanding and Exposing Performance Mismatches via Configurations.....	1712
<i>Haochen He, Erci Xu, Shanshan Li, Zhouyang Jia, Si Zheng, Yue Yu, Jun Ma, Xiangke Liao</i>	
Semantics-aware Dataset Discovery from Data Lakes with Contextualized Column-based Representation Learning .....	1726
<i>Grace Fan, Jin Wang, Yuliang Li, Dan Zhang, Renée J. Miller</i>	
Marigold: Efficient k-means Clustering in High Dimensions .....	1740
<i>Kasper Overgaard Mortensen, Fatemeh Zardbani, Mohammad Ahsanul Haque, Steinn Ymir Agustsson, Davide Mottin, Philip Hofmann, Panagiotis Karras</i>	
The Case for Learned In-Memory Joins.....	1749
<i>Ibrahim Sabek, Tim Kraska</i>	
Elf: Erasing-based Lossless Floating-Point Compression .....	1763
<i>Ruiyuan Li, Zheng Li, Yi Wu, Chao Chen, Yu Zheng</i>	
LOGGER: A Learned Optimizer towards Generating Efficient and Robust Query Execution Plans .....	1777
<i>Tianyi Chen, Jun Gao, Hedui Chen, Yaofeng Tu</i>	

Representing Paths in Graph Database Pattern Matching ..... 1790  
*Wim Martens, Matthias Niewerth, Tina Popp, Carlos Rojas, Stijn Vansummeren, Domagoj Vrgoč*



## TABLE OF CONTENTS

### Research Papers

ZKSQL: Verifiable and Efficient Query Evaluation with Zero-Knowledge Proofs.....	1804
<i>Xiling Li, Chenkai Weng, Yongxin Xu, Xiao Wang, Jennie Rogers</i>	
Computing Graph Edit Distance via Neural Graph Matching .....	1817
<i>Chengzhi Piao, Tingyang Xu, Xiangguo Sun, Yu Rong, Kangfei Zhao, Hong Cheng</i>	
Benchmarking the Utility of $w$ -event Differential Privacy Mechanisms - When Baselines Become Mighty Competitors .....	1830
<i>Christine Schäler, Thomas Hütter, Martin Schäler</i>	
Collective Grounding: Applying Database Techniques to Grounding Templated Models .....	1843
<i>Eriq Augustine, Lise Getoor</i>	
An Experimental Evaluation of Process Concept Drift Detection .....	1856
<i>Jan Niklas Adams, Cameron Pitsch, Tobias Brockhoff, Wil M. P. Van Der Aalst</i>	
Pollock: A Data Loading Benchmark .....	1870
<i>Gerardo Vitagliano, Mazhar Hameed, Lan Jiang, Lucas Reisener, Eugene Wu, Felix Naumann</i>	
Answering Private Linear Queries Adaptively using the Common Mechanism .....	1883
<i>Yingtai Xiao, Guanhong Wang, Danfeng Zhang, Daniel Kifer</i>	
LDPTrace: Locally Differentially Private Trajectory Synthesis .....	1897
<i>Yuntao Du, Yujia Hu, Zhikun Zhang, Ziquan Fang, Lu Chen, Baihua Zheng, Yunjun Gao</i>	
Sim-Piece: Highly Accurate Piecewise Linear Approximation through Similar Segment Merging.....	1910
<i>Xenophon Kitsios, Panagiotis Liakos, Katia Papakonstantinou, Yannis Kotidis</i>	
Towards Migration-Free Just-In-Case Data Archival for Future Cloud Data Lakes .....	1923
<i>Eugenio Marinelli, Yiqing Yan, Virginie Magnone, Charlotte Dumargne, Pascal Barbry, Thomas Heinis, Raja Appuswamy</i>	
Fine-Grained Re-Execution for Efficient Batched Commit of Distributed Transactions .....	1930
<i>Zhiyuan Dong, Zhaoguo Wang, Xiaodong Zhang, Xian Xu, Changgeng Zhao, Haibo Chen, Aurojit Panda, Jinyang Li</i>	
Learning and Deducing Temporal Orders.....	1944
<i>Wenfei Fan, Resul Tugay, Yaoshu Wang, Min Xie, Muhammad Asif Ali</i>	
BASE: Bridging the Gap between Cost and Latency for Query Optimization .....	1958
<i>Xu Chen, Zhen Wang, Shuncheng Liu, Yaliang Li, Kai Zeng, Bolin Ding, Jingren Zhou, Han Su, Kai Zheng</i>	
Efficient framework for operating on data sketches .....	1967
<i>Jakub Lemiesz</i>	

Towards Efficient Index Construction and Approximate Nearest Neighbor Search in High-Dimensional Spaces .....	1979
<i>Xi Zhao, Yao Tian, Kai Huang, Bolong Zheng, Xiaofang Zhou</i>	
Learned Index: A Comprehensive Experimental Evaluation.....	1992
<i>Zhaoyan Sun, Xuanhe Zhou, Guoliang Li</i>	
Longshot: Indexing Growing Databases using MPC and Differential Privacy .....	2005
<i>Yanping Zhang, Johes Bater, Kartik Nayak, Ashwin Machanavajhala</i>	
Accelerating Similarity Search for Elastic Measures: A Study and New Generalization of Lower Bounding Distances .....	2019
<i>John Paparrizos, Kaize Wu, Aaron Elmore, Christos Faloutsos, Michael J. Franklin</i>	
AdaChain: A Learned Adaptive Blockchain .....	2033
<i>Chenyuan Wu, Bhavana Mehta, Mohammad Javad Amiri, Ryan Marcus, Boon Thau Loo</i>	
Influential Community Search over Large Heterogeneous Information Networks .....	2047
<i>Yingli Zhou, Yixiang Fang, Wensheng Luo, Yunming Ye</i>	

## TABLE OF CONTENTS

### Research Papers

Neighborhood-based Hypergraph Core Decomposition .....	2061
<i>Naheed Anjum Arafat, Arijit Khan, Arpit Kumar Rai, Bishwamitra Ghosh</i>	
Temporal SIR-GN: Efficient and Effective Structural Representation Learning for Temporal Graphs.....	2075
<i>Janet Layne, Justin Carpenter, Edoardo Serra, Francesco Gullo</i>	
What Modern NVMe Storage Can Do, And How To Exploit It: High-Performance I/O for High-Performance Storage Engines .....	2090
<i>Gabriel Haas, Viktor Leis</i>	
WiscSort: External Sorting For Byte-Addressable Storage .....	2103
<i>Vinay Banakar, Kan Wu, Yuvraj Patel, Kimberly Keeton, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau</i>	
Text Indexing for Long Patterns: Anchors are All you Need .....	2117
<i>Lorraine A. K. Ayad, Grigorios Loukides, Solon P. Pissis</i>	
The FastLanes Compression Layout: Decoding >100 Billion Integers per Second with Scalar Code .....	2132
<i>Azim Afroozeh, Peter Boncz</i>	
VeriBench: Analyzing the Performance of Database Systems with Verifiability .....	2145
<i>Cong Yue, Meihui Zhang, Changhao Zhu, Gang Chen, Dumitrel Loghin, Beng Chin Ooi</i>	
Towards Designing and Learning Piecewise Space-Filling Curves .....	2158
<i>Jiangneng Li, Zheng Wang, Gao Cong, Cheng Long, Han Mao Kiah, Bin Cui</i>	
MiniGraph: Querying Big Graphs with a Single Machine .....	2172
<i>Xiaoke Zhu, Yang Liu, Shuhao Liu, Wenfei Fan</i>	
BICE: Exploring Compact Search Space by Using Bipartite Matching and Cell-Wide Verification .....	2186
<i>Yunyoung Choi, Kunsoo Park, Hyunjoon Kim</i>	
Maximal D-truss Search in Dynamic Directed Graphs .....	2199
<i>Anxin Tian, Alexander Zhou, Yue Wang, Lei Chen</i>	
DILI: A Distribution-Driven Learned Index.....	2212
<i>Pengfei Li, Hua Lu, Rong Zhu, Bolin Ding, Long Yang, Gang Pan</i>	
Pre-trained Embeddings for Entity Resolution: An Experimental Analysis .....	2225
<i>Alexandros Zeakis, George Papadakis, Dimitrios Skoutas, Manolis Koubarakis</i>	
Decoupled Graph Neural Networks for Large Dynamic Graphs.....	2239
<i>Yanping Zheng, Zhewei Wei, Jiajun Liu</i>	
Adaptive Indexing of Objects with Spatial Extent .....	2248
<i>Fatemeh Zardbani, Nikos Mamoulis, Stratos Idreos, Panagiotis Karras</i>	

LEON: A New Framework for ML-Aided Query Optimization .....	2261
<i>Xu Chen, Haitian Chen, Zibo Liang, Shuncheng Liu, Jinghong Wang, Kai Zeng, Han Su, Kai Zheng</i>	
TiQuE: Improving the Transactional Performance of Analytical Systems for True Hybrid Workloads .....	2274
<i>Nuno Faria, José Pereira, Ana Nunes Alonso, Ricardo Vilaça, Yunus Koning, Niels Nes</i>	
SEIDEN: Revisiting Query Processing in Video Database Systems .....	2289
<i>Jaeho Bang, Gaurav Tarlok Kakkar, Pramod Chunduri, Subrata Mitra, Joy Arulraj</i>	
Extract-Transform-Load for Video Streams .....	2302
<i>Ferdinand Kossmann, Ziniu Wu, Eugenie Lai, Nesime Tatbul, Lei Cao, Tim Kraska, Sam Madden</i>	
Pando: Enhanced Data Skipping with Logical Data Partitioning .....	2316
<i>Sivaprasad Sudhir, Wenbo Tao, Nikolay Laptev, Cyrille Habis, Michael Cafarella, Samuel Madden</i>	
Cracking-Like Join for Trusted Execution Environments .....	2330
<i>Kajetan Maliszewski, Jorge-Arnulfo Quiané-Ruiz, Volker Markl</i>	
Opportunities for Quantum Acceleration of Databases: Optimization of Queries and Transaction Schedules .....	2344
<i>Umut Çalikyılmaz, Sven Groppe, Jinghua Groppe, Tobias Winker, Stefan Prestel, Farida Shagieva, Daanish Arya, Florian Preis, Le Gruenwald</i>	
SDPipe: A Semi-Decentralized Framework for Heterogeneity-aware Pipeline-parallel Training .....	2354
<i>Xupeng Miao, Yining Shi, Zhi Yang, Bin Cui, Zhihao Jia</i>	
LRU-C: Parallelizing Database I/Os for Flash SSDs .....	2364
<i>Bo-Hyun Lee, Mijin An, Sang-Won Lee</i>	
Why Not Yet: Fixing a Top-k Ranking that Is Not Fair to Individuals .....	2377
<i>Zixuan Chen, Panagiotis Manolios, Mirek Riedewald</i>	

## TABLE OF CONTENTS

### Research Papers

Information-Theoretically Secure and Highly Efficient Search and Row Retrieval.....	2391
<i>Shantanu Sharma, Yin Li, Sharad Mehrotra, Nisha Panwar, Komal Kumari, Swagnik Roychoudhury</i>	
Olive: Oblivious Federated Learning on Trusted Execution Environment Against the Risk of Sparsification	2404
<i>Fumiyuki Kato, Yang Cao, Masatoshi Yoshikawa</i>	
TASK: An Efficient Framework for Instant Error-tolerant Spatial Keyword Queries on Road Networks .....	2418
<i>Chengyang Luo, Qing Liu, Yunjun Gao, Lu Chen, Ziheng Wei, Congcong Ge</i>	
Autonomously Computable Information Extraction.....	2431
<i>Besat Kassaie, Frank Wm. Tompa</i>	
NVM: Is it Not Very Meaningful for Databases?.....	2444
<i>Dimitrios Koutsoukos, Raghav Bhartia, Michal Friedman, Ana Klimovic, Gustavo Alonso</i>	
DeepJoin: Joinable Table Discovery with Pre-trained Language Models .....	2458
<i>Yuyang Dong, Chuan Xiao, Takuma Nozawa, Masafumi Enomoto, Masafumi Oyamada</i>	
Falcon: A Privacy-Preserving and Interpretable Vertical Federated Learning System.....	2471
<i>Yuncheng Wu, Naili Xing, Gang Chen, Tien Tuan Anh Dinh, Zhaojing Luo, Beng Chin Ooi, Xiaokui Xiao, Meihui Zhang</i>	
Enabling Secure and Efficient Data Analytics Pipeline Evolution with Trusted Execution Environment.....	2485
<i>Haotian Gao, Cong Yue, Tien Tuan Anh Dinh, Zhiyong Huang, Beng Chin Ooi</i>	
A Case for Graphics-driven Query Processing .....	2499
<i>Harish Doraiswamy, Vikas Kalagi, Karthik Ramachandra, Jayant R. Haritsa</i>	
Effective and Efficient Route Planning Using Historical Trajectories on Road Networks .....	2512
<i>Wei Tian, Jieming Shi, Siqiang Luo, Hui Li, Xike Xie, Yuanhang Zou</i>	
Adaptive Indexing in High-Dimensional Metric Spaces.....	2525
<i>Konstantinos Lampropoulos, Fatemeh Zardbani, Nikos Mamoulis, Panagiotis Karras</i>	
Parallel Colorful h-star Core Maintenance in Dynamic Graphs.....	2538
<i>Sen Gao, Hongchao Qin, Rong-Hua Li, Bingsheng He</i>	
MITra: A Framework for Multi-Instance Graph Traversal .....	2551
<i>Jia Li, Wenyue Zhao, Nikos Ntarmos, Yang Cao, Peter Buneman</i>	
CommunityAF: An Example-based Community Search Method via Autoregressive Flow.....	2565
<i>Jiazun Chen, Yikuan Xia, Jun Gao</i>	
Auto-BI: Automatically Build BI-Models Leveraging Local Join Prediction and Global Schema Graph .....	2578
<i>Yiming Lin, Yeye He, Surajit Chaudhuri</i>	

Trajectory Data Collection with Local Differential Privacy .....	2591
<i>Yuemin Zhang, Qingqing Ye, Rui Chen, Haibo Hu, Qilong Han</i>	
LMSFC: A Novel Multidimensional Index based on Learned Monotonic Space Filling Curves.....	2605
<i>Jian Gao, Xin Cao, Xin Yao, Gong Zhang, Wei Wang</i>	
Scaling a Declarative Cluster Manager Architecture with Query Optimization Techniques.....	2618
<i>Xexin Rong, Mihai Budiu, Athinagoras Skiadopoulos, Lalith Suresh, Amy Tai</i>	
CORNET: Learning Table Formatting Rules By Example .....	2632
<i>Mukul Singh, José Cambroneró Sánchez, Sumit Gulwani, Vu Le, Carina Negreanu, Mohammad Raza, Gust Verbruggen</i>	
ARKGraph: All-Range Approximate K-Nearest-Neighbor Graph.....	2645
<i>Chaoji Zuo, Dong Deng</i>	
Causal Data Integration.....	2659
<i>Brit Youngmann, Michael Cafarella, Babak Salimi, Anna Zeng</i>	
Mining Frequent Infix Patterns from Concurrency-Aware Process Execution Variants.....	2666
<i>Michael Martini, Daniel Schuster, Wil M.P. Van Der Aalst</i>	
The Composable Data Management System Manifesto .....	2679
<i>Pedro Pedreira, Orri Erling, Konstantinos Karanasos, Scott Schneider, Wes Mc Kinney, Satya R Valluri, Mohamed Zait, Jacques Nadeau</i>	

## TABLE OF CONTENTS

### Research Papers

A Two-Level Signature Scheme for Stable Set Similarity Joins.....	2686
<i>Daniel Schmitt, Daniel Kocher, Nikolaus Augsten, Willi Mann, Alexander Miller</i>	
Scalable Reasoning on Document Stores via Instance-Aware Query Rewriting.....	2699
<i>Olivier Rodriguez, Federico Ulliana, Marie-Laure Mugnier</i>	
EQUI-VOCAL: Synthesizing Queries for Compositional Video Events from Limited User Interactions.....	2714
<i>Enhao Zhang, Maureen Daum, Dong He, Brandon Haynes, Ranjay Krishna, Magdalena Balazinska</i>	
Lotan: Bridging the Gap between GNNs and Scalable Graph Analytics Engines.....	2728
<i>Yuhao Zhang, Arun Kumar</i>	
Epoxy: ACID Transactions Across Diverse Data Stores.....	2742
<i>Peter Kraft, Qian Li, Xinjing Zhou, Peter Bailis, Michael Stonebraker, Xiangyao Yu, Matei Zaharia</i>	
Analyzing Vectorized Hash Tables Across CPU Architectures.....	2755
<i>Maximilian Böther, Lawrence Benson, Ana Klimovic, Tilmann Rabl</i>	
Exploiting Cloud Object Storage for High-Performance Analytics.....	2769
<i>Dominik Durner, Viktor Leis, Thomas Neumann</i>	
A Randomized Blocking Structure for Streaming Record Linkage.....	2783
<i>Dimitrios Karapiperis, Christos Tjortjis, Vassilios S. Verykios</i>	
REmatch: a novel regex engine for finding all matches.....	2792
<i>Cristian Riveros, Nicolás Van Sint Jan, Domagoj Vrgoč</i>	
ADOPT: Adaptively Optimizing Attribute Orders for Worst-Case Optimal Join Algorithms via Reinforcement Learning.....	2805
<i>Junxiong Wang, Immanuel Trummer, Ahmet Kara, Dan Olteanu</i>	
Triangular Stability Maximization by Influence Spread over Social Networks.....	2818
<i>Zheng Hu, Weiguo Zheng, Xiang Lian</i>	
CORE-Sketch: On Exact Computation of Median Absolute Deviation with Limited Space.....	2832
<i>Haoquan Guan, Ziling Chen, Shaoxu Song</i>	
Fast Search-By-Classification for Large-Scale Databases Using Index-Aware Decision Trees and Random Forests.....	2845
<i>Christian Lülfi, Denis Mayr Lima Martins, Marcos Antonio Vaz Salles, Yongluan Zhou, Fabian Gieseke</i>	
Semi-Oblivious Chase Termination for Linear Existential Rules: An Experimental Study.....	2858
<i>Marco Calautti, Mostafa Milani, Andreas Pieris</i>	
Analyzing the Impact of Cardinality Estimation on Execution Plans in Microsoft SQL Server.....	2871
<i>Kukjin Lee, Anshuman Dutt, Vivek Narasayya, Surajit Chaudhuri</i>	

WALTZ: Leveraging Zone Append to Tighten the Tail Latency of LSM Tree on ZNS SSD .....	2884
<i>Jongsung Lee, Donguk Kim, Jae W. Lee</i>	
Accelerating Aggregation Queries on Unstructured Streams of Data .....	2897
<i>Matthew Russo, Tatsunori Hashimoto, Daniel Kang, Yi Sun, Matei Zaharia</i>	
QueryBooster: Improving SQL Performance Using Middleware Services for Human-Centered Query Rewriting .....	2911
<i>Qiushi Bai, Sadeem Alsudais, Chen Li</i>	
Consistent Range Approximation for Fair Predictive Modeling .....	2925
<i>Jiongli Zhu, Sainyam Galhotra, Nazanin Sabri, Babak Salimi</i>	
SUREL+: Moving from Walks to Sets for Scalable Subgraph-based Graph Representation Learning .....	2939
<i>Haoteng Yin, Muhan Zhang, Jianguo Wang, Pan Li</i>	
Estimating Single-Node PageRank in $\tilde{O}(\min\{d_t, \sqrt{m}\})$ Time .....	2949
<i>Hanzhi Wang, Zhewei Wei</i>	
Simple Adaptive Query Processing vs. Learned Query Optimizers: Observations and Analysis.....	2962
<i>Yunjia Zhang, Yannis Chronis, Jignesh M. Patel, Theodoros Rekatsinas</i>	
BP-tree: Overcoming the Point-Range Operation Tradeoff for In-Memory B-trees .....	2976
<i>Helen Xu, Amanda Li, Brian Wheatman, Manoj Marneni, Prashant Pandey</i>	
HENCE-X: Toward Heterogeneity-agnostic Multi-level Explainability for Deep Graph Networks .....	2990
<i>Ge Lv, Chen Jason Zhang, Lei Chen</i>	
Automatic Road Extraction with Multi-Source Data Revisited: Completeness, Smoothness and Discrimination .....	3004
<i>Haitao Yuan, Sai Wang, Zhifeng Bao, Shangguang Wang</i>	
Asymptotically Better Query Optimization Using Indexed Algebra.....	3018
<i>Philipp Fent, Guido Moerkotte, Thomas Neumann</i>	
Normalizing Property Graphs .....	3031
<i>Philipp Skavantzios, Sebastian Link</i>	
A Deep Dive into Common Open Formats for Analytical DBMSs .....	3044
<i>Chunwei Liu, Anna Pavlenko, Matteo Interlandi, Brandon Haynes</i>	
Saibot: A Differentially Private Data Search Platform.....	3057
<i>Zezhou Huang, Jiaxiang Liu, Daniel Gbenga Alabi, Raul Castro Fernandez, Eugene Wu</i>	
JoinBoost: Grow Trees Over Normalized Data Using Only SQL.....	3071
<i>Zezhou Huang, Rathijit Sen, Jiaxiang Liu, Eugene Wu</i>	
R <sup>3</sup> : Record-Replay-Retroaction for Database-Backed Applications.....	3085
<i>Qian Li, Peter Kraft, Michael Cafarella, Çağatay Demiralp, Goetz Graefe, Christos Kozyrakis, Michael Stonebraker, Lalith Suresh, Xiangyao Yu, Matei Zaharia</i>	
Self-Training for Label-Efficient Information Extraction from Semi-Structured Web-Pages .....	3098
<i>Ritesh Sarkhel, Binxuan Huang, Colin Lockard, Prashant Shiralkar</i>	
Efficient Non-Learning Similar Subtrajectory Search.....	3111
<i>Jiabao Jin, Peng Cheng, Lei Chen, Xuemin Lin, Wenjie Zhang</i>	



Frequency-revealing attacks against Frequency-hiding Order-preserving Encryption .....	3124
<i>Xinle Cao, Jian Liu, Yongsheng Shen, Xiaohua Ye, Kui Ren</i>	
Efficient Fault Tolerance for Recommendation Model Training via Erasure Coding .....	3137
<i>Tianyu Zhang, Kaige Liu, Jack Kosaian, Juncheng Yang, Rashmi Vinayak</i>	
SlabCity: Whole-Query Optimization using Program Synthesis .....	3151
<i>Rui Dong, Jie Liu, Yuxuan Zhu, Cong Yan, Barzan Mozafari, Xinyu Wang</i>	
Scaling Up Structural Clustering to Large Probabilistic Graphs Using Lyapunov Central Limit Theorem.....	3165
<i>Joseph Howie, Venkatesh Srinivasan, Alex Thomo</i>	
Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy .....	3178
<i>Lucas Rosenblatt, Bernease Herman, Anastasia Holovenko, Wonkwon Lee, Joshua Loftus, Elizabeth Mckinnie, Taras Rumezhak, Andrii Stadnik, Bill Howe, Julia Stoyanovich</i>	
POEM: Pattern-Oriented Explanations of Convolutional Neural Networks .....	3192
<i>Vargha Dadvar, Lukasz Golab, Divesh Srivastava</i>	
gCore: Exploring Cross-layer Cohesiveness in Multi-layer Graphs.....	3201
<i>Dandan Liu, Zhaonian Zou</i>	
Sieve: A Learned Data-Skipping Index for Data Analytics.....	3214
<i>Yulai Tong, Jiazhen Liu, Hua Wang, Ke Zhou, Rongfeng He, Qin Zhang, Cheng Wang</i>	
Out-of-Order Sliding-Window Aggregation with Efficient Bulk Evictions and Insertions.....	3227
<i>Kanat Tangwongsan, Martin Hirzel, Scott Schneider</i>	
k-Best Egalitarian Stable Marriages for Task Assignment .....	3240
<i>Siyuan Wu, Leong Hou U, Panagiotis Karras</i>	
Federated Calibration and Evaluation of Binary Classifiers.....	3253
<i>Graham Cormode, Igor L. Markov</i>	
FlashAlloc: Dedicating Flash Blocks By Objects .....	3266
<i>Jonghyeok Park, Soyee Choi, Gihwan Oh, Soojun Im, Moon-Wook Oh, Sang-Won Lee</i>	
Through the Fairness Lens: Experimental Analysis and Evaluation of Entity Matching.....	3279
<i>Nima Shahbazi, Nikola Danevski, Fatemeh Nargesian, Abolfazl Asudeh, Divesh Srivastava</i>	
Check Out the Big Brain on BRAD: Simplifying Cloud Data Processing with Learned Automated Data Meshes .....	3293
<i>Tim Kraska, Tianyu Li, Samuel Madden, Markos Markakis, Amadou Ngom, Ziniu Wu, Geoffrey X. Yu</i>	
How Large Language Models Will Disrupt Data Management.....	3302
<i>Raul Castro Fernandez, Aaron J. Elmore, Michael J. Franklin, Sanjay Krishnan, Chenhao Tan</i>	
FASTgres: Making Learned Query Optimizer Hinting Effective .....	3310
<i>Lucas Woltmann, Jerome Thiessat, Claudio Hartmann, Dirk Habich, Wolfgang Lehner</i>	
Write-Aware Timestamp Tracking: Effective and Efficient Page Replacement for Modern Hardware .....	3323
<i>Demian Vöhringer, Viktor Leis</i>	

Tigger: A Database Proxy That Bounces With User-Bypass.....	3335
<i>Matthew Butrovich, Karthik Ramanathan, John Rollinson, Wan Shen Lim, William Zhang, Justine Sherry, Andrew Pavlo</i>	
Equitable Data Valuation Meets the Right to Be Forgotten in Model Markets.....	3349
<i>Haocheng Xia, Jinfei Liu, Jian Lou, Zhan Qin, Kui Ren, Yang Cao, Li Xiong</i>	
TSM-Bench: Benchmarking Time Series Database Systems for Monitoring Applications.....	3363
<i>Abdelouahab Khelifati, Mourad Khayati, Anton Dignös, Djellel Difallah, Philippe Cudré-Mauroux</i>	
Cross Modal Data Discovery over Structured and Unstructured Data Lakes .....	3377
<i>Mohamed Y. Eltabakh, Mayuresh Kunjir, Ahmed K. Elmagarmid, Mohammad Shahmeer Ahmad</i>	
Auto-Tables: Synthesizing Multi-Step Transformations to Relationalize Tables without Using Examples .....	3391
<i>Peng Li, Yeye He, Cong Yan, Yue Wang, Surajit Chaudhuri</i>	
Effective Entity Augmentation By Querying External Data Sources.....	3404
<i>Christopher Buss, Jasmin Mousavi, Mikhail Tokarev, Arash Termehchy, David Maier, Stefan Lee</i>	
Choose Wisely: An Extensive Evaluation of Model Selection for Anomaly Detection in Time Series.....	3418
<i>Emmanouil Sylligardos, Paul Boniol, John Paparrizos, Panos Trahanias, Themis Palpanas</i>	
Similarity search in the blink of an eye with compressed indices .....	3433
<i>Cecilia Aguerrebere, Ishwar Singh Bhati, Mark Hildebrand, Mariano Tepper, Theodore Willke</i>	
On Data-Aware Global Explainability of Graph Neural Networks .....	3447
<i>Ge Lv, Lei Chen</i>	
Declarative Sub-Operators for Universal Data Processing.....	3461
<i>Michael Jungmair, Jana Giceva</i>	

## TABLE OF CONTENTS

### Industrial Papers

Progressive Partitioning for Parallelized Query Execution in Google's Napa .....	3475
<i>Junichi Tatemura, Tao Zou, Jagan Sankaranarayanan, Yanlai Huang, Jim Chen, Yupu Zhang, Kevin Lai, Hao Zhang, Gokul Nath Babu Manoharan, Goetz Graefe, Divyakant Agrawal, Brad Adelberg, Shilpa Kolhar, Indrajit Roy</i>	
Taurus MM: bringing multi-master to the cloud .....	3488
<i>Alex Depoutovitch, Chong Chen, Per-Åke Larson, Jack Ng, Shu Lin, Guanzhu Xiong, Paul Lee, Emad Boctor, Samiao Ren, Lengdong Wu, Yuchen Zhang, Calvin Sun</i>	
StreamOps: Cloud-Native Runtime Management for Streaming Services in ByteDance.....	3501
<i>Yancan Mao, Zhanghao Chen, Yifan Zhang, Meng Wang, Yong Fang, Guanghui Zhang, Rui Shi, Richard T.B. Ma</i>	
AutoSteer: Learned Query Optimization for Any SQL Database.....	3515
<i>Christoph Anneser, Nesime Tatbul, David Cohen, Zhenggang Xu, Prithviraj Pandian, Nikolay Laptev, Ryan Marcus</i>	
Krypton: Real-time Serving and Analytical SQL Engine at ByteDance.....	3528
<i>Jianjun Chen, Rui Shi, Heng Chen, Li Zhang, Ruidong Li, Wei Ding, Liya Fan, Hao Wang, Mu Xiong, Yuxiang Chen, Benchao Dong, Kuankuan Guo, Yuanjin Lin, Xiao Liu, Haiyang Shi, Peipei Wang, Zikang Wang, Yemeng Yang, Junda Zhao, Dongyan Zhou, Zhikai Zuo, Yuming Liang</i>	
EmbedX: A Versatile, Efficient and Scalable Platform to Embed Both Graphs and High-Dimensional Sparse Data.....	3543
<i>Yuanhang Zou, Zhihao Ding, Jieming Shi, Shuting Guo, Chunchen Su, Yafei Zhang</i>	
The Story of AWS Glue.....	3557
<i>Mohit Saxena, Benjamin Sowell, Daiyan Alamgir, Nitin Bahadur, Bijay Bisht, Santosh Chandrachood, Chitti Keswani, G2 Krishnamoorthy, Austin Lee, Bohou Li, Zach Mitchell, Vaibhav Porwal, Maheedhar Reddy Chappidi, Brian Ross, Noritaka Sekiyama, Omer Zaki, Linchi Zhang, Mehul A. Shah</i>	
Towards General and Efficient Online Tuning for Spark.....	3570
<i>Yang Li, Huaijun Jiang, Yu Shen, Yide Fang, Xiaofeng Yang, Danqing Huang, Xinyi Zhang, Wentao Zhang, Ce Zhang, Peng Chen, Bin Cui</i>	
CDSBen: Benchmarking the Performance of Storage Services in Cloud-native Database System at ByteDance .....	3584
<i>Jiashu Zhang, Wen Jiang, Bo Tang, Haoxiang Ma, Lixun Cao, Zhongbin Jiang, Yuanyuan Nie, Fan Wang, Lei Zhang, Yuming Liang</i>	
FEBench: A Benchmark for Real-Time Relational Data Feature Extraction.....	3597
<i>Xuanhe Zhou, Cheng Chen, Kunyi Li, Bingsheng He, Mian Lu, Qiaosheng Liu, Wei Huang, Guoliang Li, Zhao Zheng, Yuqiang Chen</i>	

MINT: Detecting Fraudulent Behaviors from Time-series Relational Data.....	3610
<i>Fei Xiao, Yuncheng Wu, Meihui Zhang, Gang Chen, Beng Chin Ooi</i>	
Microsoft Purview: A System for Central Governance of Data .....	3624
<i>Shafi Ahmad, Dillidorai Arumugam, Srdan Bozovic, Elnata Degefa, Sailesh Duvvuri, Steven Gott, Nitish Gupta, Joachim Hammer, Nivedita Kaluskar, Raghav Kaushik, Rakesh Khanduja, Prasad Mujumdar, Gaurav Malhotra, Pankaj Naik, Nikolas Ogg, Krishna Kumar Parthasarthy, Raghu Ramakrishnan, Vlad Rodriguez, Rahul Sharma, Jakub Szymaszek, Andreas Wolter</i>	
Anser: Adaptive Information Sharing Framework of AnalyticDB .....	3636
<i>Liang Lin, Yuhan Li, Bin Wu, Huijun Mai, Renjie Lou, Jian Tan, Feifei Li</i>	
TPCx-AI - An Industry Standard Benchmark for Artificial Intelligence and Machine Learning Systems.....	3649
<i>Christoph Brücke, Philipp Härtling, Rodrigo D Escobar Palacios, Hamesh Patel, Tilmann Rabl</i>	
OneProvenance: Efficient Extraction of Dynamic Coarse-Grained Provenance From Database Query Event Logs.....	3662
<i>Fotis Psallidas, Ashvin Agrawal, Chandru Sugunan, Khaled Ibrahim, Konstantinos Karanasos, Jesús Camacho-Rodríguez, Avrielia Floratou, Carlo Curino, Raghu Ramakrishnan</i>	
Techniques and Efficiencies from Building a Real-Time DBMS .....	3676
<i>V. Srinivasan, Andrew Gooding, Sunil Sayyaparaju, Thomas Lopatic, Kevin Porter, Ashish Shinde, B. Narendran</i>	
Real-time Workload Pattern Analysis for Large-scale Cloud Databases .....	3689
<i>Jiaqi Wang, Tianyi Li, Anni Wang, Xiaozhe Liu, Lu Chen, Jie Chen, Jianye Liu, Junyang Wu, Feifei Li, Yunjun Gao</i>	
Big Data Analytic Toolkit: A general-purpose, modular, and heterogeneous acceleration toolkit for data analytical engines .....	3702
<i>Jiang Li, Qi Xie, Yan Ma, Jian Ma, Kunshang Ji, Yizhong Zhang, Chaojun Zhang, Yixiu Chen, Gangsheng Wu, Jie Zhang, Kaidi Yang, Xinyi He, Qiuyang Shen, Yanting Tao, Haiwei Zhao, Penghui Jiao, Chengfei Zhu, David Qian, Cheng Xu</i>	
Lindorm TSDB: A Cloud-native Time-series Database for Large-scale Monitoring Systems.....	3715
<i>Chunhui Shen, Qianyu Ouyang, Feibo Li, Zhipeng Liu, Longcheng Zhu, Yujie Zou, Qing Su, Tianhuan Yu, Yi Yi, Jianhong Hu, Cen Zheng, Bo Wen, Hanbang Zheng, Lunfan Xu, Sicheng Pan, Bin Wu, Xiao He, Ye Li, Jian Tan, Sheng Wang, Dan Pei, Wei Zhang, Feifei Li</i>	
OceanBase Paetica: A Hybrid Shared-nothing/Shared-everything Database for Supporting Single Machine and Distributed Cluster .....	3728
<i>Zhifeng Yang, Quanqing Xu, Shanyan Gao, Chuanhui Yang, Guoping Wang, Yuzhong Zhao, Fanyu Kong, Hao Liu, Wanhong Wang, Jinliang Xiao</i>	
SimpleTS: An Efficient and Universal Model Selection Framework for Time Series Forecasting .....	3741
<i>Yuanyuan Yao, Dimeng Li, Hailiang Jie, Lu Chen, Tianyi Li, Jie Chen, Jiaqi Wang, Feifei Li, Yunjun Gao</i>	
PolarDB-SCC: A Cloud-Native Database Ensuring Low Latency for Strongly Consistent Reads .....	3754
<i>Xinjun Yang, Yingqiang Zhang, Hao Chen, Chuan Sun, Feifei Li, Wenchao Zhou</i>	
ScalarDB: Universal Transaction Manager for Polystores.....	3768
<i>Hirokyu Yamada, Toshihiro Suzuki, Yuji Ito, Jun Nemoto</i>	
Angel-PTM: A Scalable and Economical Large-scale Pre-training System in Tencent .....	3781
<i>Xiaonan Nie, Yi Liu, Fangcheng Fu, Jinbao Xue, Dian Jiao, Xupeng Miao, Yangyu Tao, Bin Cui</i>	
Eigen: End-to-end Resource Optimization for Large-Scale Databases on the Cloud .....	3795

*Ji You Li, Jiachi Zhang, Wenchao Zhou, Yuhang Liu, Shuai Zhang, Zhuoming Xue, Ding Xu, Hua Fan, Fangyuan Zhou, Feifei Li*

MagicScaler: Uncertainty-aware, Predictive Autoscaling..... 3808  
*Zhicheng Pan, Yihang Wang, Yingying Zhang, Sean Bin Yang, Yunyao Cheng, Peng Chen, Chenjuan Guo, Qingsong Wen, Xiduo Tian, Yunliang Dou, Zhiqiang Zhou, Chengcheng Yang, Aoying Zhou, Bin Yang*

Kora: A Cloud-Native Event Streaming Platform for Kafka..... 3822  
*Anna Povzner, Prince Mahajan, Jason Gustafson, Jun Rao, Ismael Juma, Feng Min, Shriram Sridharan, Nikhil Bhatia, Gopi Attaluri, Adithya Chandra, Stanislav Kozlovski, Rajini Sivaram, Lucas Bradstreet, Bob Barrett, Dhruvil Shah, David Jacot, David Arthur, Manveer Chawla, Ron Dagostino, Colin Mccabe, Manikumar Reddy Obili, Kowshik Prakasam, Jose Garcia Sancio, Vikas Singh, Alok Nikhil, Kamal Gupta*

Automatic SQL Error Mitigation in Oracle ..... 3835  
*Krishna Kantikiran Pasupuleti, Jiakun Li, Hong Su, Mohamed Ziauddin*

PyTorch FSDP: Experiences on Scaling Fully Sharded Data Parallel..... 3848  
*Yanli Zhao, Andrew Gu, Rohan Varma, Liang Luo, Chien-Chin Huang, Min Xu, Less Wright, Hamid Shojanazeri, Myle Ott, Sam Shleifer, Alban Desmaison, Can Balioglu, Pritam Damania, Bernard Nguyen, Geeta Chauhan, Yuchen Hao, Ajit Mathews, Shen Li*

## **Tutorials**

Time Series Data Mining: A Unifying View..... 3861  
*Eamonn Keogh*

Machine Learning for Subgraph Extraction: Methods, Applications and Challenges ..... 3864  
*Kai Siong Yow, Ningyi Liao, Siqiang Luo, Reynold Cheng*

Private Information Retrieval in Large Scale Public Data Repositories..... 3868  
*Ishtiyaque Ahmad, Divyakant Agrawal, Amr El Abbadi, Trinabh Gupta*

Data and AI Model Markets: Opportunities for Data and Model Sharing, Discovery, and Integration..... 3872  
*Jian Pei, Raul Castro Fernandez, Xiaohui Yu*

Efficient Execution of User-Defined Functions in SQL Queries..... 3874  
*Yannis Foufoulas, Alkis Simitsis*

Natural Language Interfaces for Databases with Deep Learning..... 3878  
*George Katsogiannis-Meimarakis, Mike Xydas, Georgia Koutrika*

Join Order Selection with Deep Reinforcement Learning: Fundamentals, Techniques, and Challenges ..... 3882  
*Zhengdong Yan, Valter Uotila, Jiaheng Lu*

Full-Power Graph Querying: State of the Art and Challenges ..... 3886  
*Ioana Manolescu, Madhulika Mohanty*

A Tutorial on Visual Representations of Relational Queries..... 3890  
*Wolfgang Gatterbauer*

Databases on Modern Networks: A Decade of Research that now comes into Practice ..... 3894  
*Alberto Lerner, Carsten Binnig, Philippe Cudré-Mauroux, Rana Hussein, Matthias Jasný, Theo Jepsen, Dan R. K. Ports, Lasse Thostrup, Tobias Ziegler*

Building a Collaborative Data Analytics System: Opportunities and Challenges..... 3898  
*Zuozhi Wang, Chen Li*

## Demonstrations

PAINE Demo: Optimizing Video Selection Queries With Commonsense Knowledge .....	3902
<i>Wenjia He, Ibrahim Sabek, Yuze Lou, Michael Cafarella</i>	
DoveDB: A Declarative and Low-Latency Video Database.....	3906
<i>Ziyang Xiao, Dongxiang Zhang, Zepeng Li, Sai Wu, Kian-Lee Tan, Gang Chen</i>	
DeepVQL: Deep Video Queries on PostgreSQL .....	3910
<i>Dong June Lew, Kihyun Yoo, Kwang Woo Nam</i>	
A Demonstration of DLBD: Database Logic Bug Detection System .....	3914
<i>Xiu Tang, Sai Wu, Dongxiang Zhang, Ziyue Wang, Gongsheng Yuan, Gang Chen</i>	
PSFQ: A Blockchain-based Privacy-preserving and Verifiable Student Feedback Questionnaire Platform .....	3918
<i>Wangze Ni, Pengze Chen, Lei Chen</i>	
QO-Insight: Inspecting Steered Query Optimizers .....	3922
<i>Christoph Anneser, Mario Petruccelli, Nesime Tatbul, David Cohen, Zhenggang Xu, Prithviraj Pandian, Nikolay Laptev, Ryan Marcus, Alfons Kemper</i>	
Lynx: A Graph Query Framework for Multiple Heterogeneous Data Sources.....	3926
<i>Zhihong Shen, Chuan Hu, Zihao Zhao</i>	
Showcasing Data Management Challenges for Future IoT Applications with NebulaStream.....	3930
<i>Aljoscha Lepping, Hoang Mi Pham, Laura Mons, Balint Rueb, Philipp M. Grulich, Ankit Chaudhary, Steffen Zeuch, Volker Markl</i>	
CEDA: Learned Cardinality Estimation with Domain Adaptation.....	3934
<i>Zilong Wang, Qixiong Zeng, Ning Wang, Haowen Lu, Yue Zhang</i>	
FastMosaic in Action: A New Mosaic Operator for Array DBMSs.....	3938
<i>Ramon Antonio Rodrigues Zalipynis</i>	
Sniffer: A Novel Model Type Detection System against Machine-Learning-as-a-Service Platforms .....	3942
<i>Zhuo Ma, Yilong Yang, Bin Xiao, Yang Liu, Xinjing Liu, Zhuoran Ma, Tong Yang</i>	
KGNav: A Knowledge Graph Navigational Visual Query System.....	3946
<i>Xiang Wang, Xin Wang, Zhaozhuo Li, Dong Han</i>	
On-the-fly Data Transformation in Action .....	3950
<i>Ju Hyoung Mun, Konstantinos Karatsenidis, Tarikul Islam Papon, Shahin Roozkhosh, Denis Hoornaert, Ulrich Drepper, Ahmed Sanaullah, Renato Mancuso, Manos Athanassoulis</i>	
Demonstrating Waffle: A Self-driving Grid Index.....	3954
<i>Dalsu Choi, Hyunsik Yoon, Hyubjin Lee, Yon Dohn Chung</i>	
CM-Explorer: Dissecting Data Ingestion Problems .....	3958
<i>Niels Bylois, Frank Neven, Stijn Vansummeren</i>	
Explaining Differentially Private Query Results With DPXplain .....	3962
<i>Tingyu Wang, Yuchao Tao, Amir Gilad, Ashwin Machanavajjhala, Sudeepa Roy</i>	
Ganos Aero: A Cloud-Native System for Big Raster Data Management and Processing.....	3966
<i>Fei Xiao, Jiong Xie, Zhida Chen, Feifei Li, Zhen Chen, Jianwei Liu, Yinpei Liu</i>	

Demonstration of OpenDBML, a Framework for Democratizing In-Database Machine Learning.....	3970
<i>Mahdi Ghorbani, Amir Shaikhha</i>	
Demonstration of SPARQL-ML: An Interfacing Language for Supporting Graph Machine Learning for RDF Graphs .....	3974
<i>Hussein Abdallah, Waleed Afandi, Essam Mansour</i>	
EQUI-VOCAL Demonstration: Synthesizing Video Queries from User Interactions.....	3978
<i>Enhao Zhang, Maureen Daum, Dong He, Manasi Ganti, Brandon Haynes, Ranjay Krishna, Magdalena Balazinska</i>	
TsQuality: Measuring Time Series Data Quality in Apache IoTDB .....	3982
<i>Yuanhui Qiu, Chenguang Fang, Shaoxu Song, Xiangdong Huang, Chen Wang, Jianmin Wang</i>	
Approximate Queries over Concurrent Updates.....	3986
<i>Congying Wang, Nithin Sastry Tellapuri, Sphoorthi Keshannagari, Dylan Zinsley, Zhuoyue Zhao, Dong Xie</i>	
Solving Hard Variants of Database Schema Matching on Quantum Computers.....	3990
<i>Kristin Fritsch, Stefanie Scherzinger</i>	
Interpretable Clustering of Multivariate Time Series with Time2Feat .....	3994
<i>Angela Bonifati, Francesco Del Buono, Francesco Guerra, Miki Lombardi, Donato Tiano</i>	
DHive: Query Execution Performance Analysis via Dataflow in Apache Hive .....	3998
<i>Chaozu Zhang, Qiaomu Shen, Bo Tang</i>	
MLWHATIF: What If You Could Stop Re-Implementing Your Machine Learning Pipeline Analyses Over and Over? .....	4002
<i>Stefan Grafberger, Shubha Guha, Paul Groth, Sebastian Schelter</i>	
AQUA: Automatic Collaborative Query Processing in Analytical Database .....	4006
<i>Yuchen Peng, Ke Chen, Lidan Shou, Dawei Jiang, Gang Chen</i>	
VisualNeo: Bridging the Gap between Visual Query Interfaces and Graph Query Engines.....	4010
<i>Kai Huang, Houdong Liang, Chongchong Yao, Xi Zhao, Yue Cui, Yao Tian, Ruiyuan Zhang, Xiaofang Zhou</i>	
KG-Roar: Interactive Datalog-based Reasoning on Virtual Knowledge Graphs .....	4014
<i>Luigi Bellomarini, Marco Benedetti, Andrea Gentili, Davide Magnanimi, Emanuel Sallinger</i>	
To UDFs and Beyond: Demonstration of a Fully Decomposed Data Processor for General Data Wrangling Tasks .....	4018
<i>Nico Schäfer, Damjan Gjurovski, Angjela Davitkova, Sebastian Michel</i>	
ChainDash: An Ad-Hoc Blockchain Data Analytics System.....	4022
<i>Yushi Liu, Liwei Yuan, Zhihao Chen, Yekai Yu, Zhao Zhang, Cheqing Jin, Ying Yan</i>	
BrewER: Entity Resolution On-Demand .....	4026
<i>Luca Zecchini, Giovanni Simonini, Sonia Bergamaschi, Felix Naumann</i>	
Visualizing Spreadsheet Formula Graphs Compactly.....	4030
<i>Fanchao Chen, Dixin Tang, Haotian Li, Aditya G. Parameswaran</i>	
DuckPGQ: Bringing SQL/PGQ to DuckDB.....	4034
<i>Daniel Ten Wolde, Gábor Szárnyas, Peter Boncz</i>	

Demo of QueryBooster: Supporting Middleware-based SQL Query Rewriting as a Service.....	4038
<i>Qiushi Bai, Sadeem Alsudais, Chen Li</i>	
Web Connector: A Unified API Wrapper to Simplify Web Data Collection.....	4042
<i>Weiyuan Wu, Pei Wang, Yi Xie, Yejia Liu, George Chow, Jiannan Wang</i>	
FS-Real: A Real-World Cross-Device Federated Learning Platform.....	4046
<i>Dawei Gao, Daoyuan Chen, Zitao Li, Yuexiang Xie, Xuchen Pan, Yaliang Li, Bolin Ding, Jingren Zhou</i>	
ADOps: An Anomaly Detection Pipeline in Structured Logs.....	4050
<i>Xintong Song, Yusen Zhu, Jianfei Wu, Bai Liu, Hongkang Wei</i>	
Portals: A Showcase of Multi-Dataflow Stateful Serverless.....	4054
<i>Jonas Spenger, Chengyang Huang, Philipp Haller, Paris Carbone</i>	
CORNET: Learning Spreadsheet Formatting Rules By Example.....	4058
<i>Mukul Singh, José Cambroner Sanchez, Sumit Gulwani, Vu Le, Carina Negreanu, Gust Verbruggen</i>	
Fanglue: An Interactive System for Decision Rule Crafting.....	4062
<i>Chen Qian, Shiwei Liang, Zhaoyang Wang, Yin Lou</i>	
Odyssey: An Engine Enabling The Time-Series Clustering Journey.....	4066
<i>John Paparrizos, Sai Prasanna Teja Reddy</i>	
ERICA: Query Refinement for Diversity Constraint Satisfaction.....	4070
<i>Jinyang Li, Alon Silberstein, Yuval Moskovitch, Julia Stoyanovich, H. V. Jagadish</i>	
Lingua Manga: A Generic Large Language Model Centric System for Data Curation.....	4074
<i>Zui Chen, Lei Cao, Sam Madden</i>	
XDB in Action: Decentralized Cross-Database Query Processing for Black-Box DBMSes.....	4078
<i>Haralampos Gavriilidis, Leonhard Rose, Joel Ziegler, Kaustubh Beedkar, Jorge-Arnulfo Quiané-Ruiz, Volker Markl</i>	
Interactive Demonstration of EVA.....	4082
<i>Gaurav Tarlok Kakkar, Aryan Rajoria, Myna Prasanna Kalluraya, Ashmita Raju, Jiashen Cao, Kexin Rong, Joy Arulraj</i>	
RESCU-SQL: Oblivious Querying for the Zero Trust Cloud.....	4086
<i>Xiling Li, Gefei Tan, Xiao Wang, Jennie Rogers, Soamar Homs</i>	
DataRinse: Semantic Transforms for Data preparation based on Code Mining.....	4090
<i>Ibrahim Abdelaziz, Julian Dolby, Udayan Khurana, Horst Samulowitz, Kavitha Srinivas</i>	
Demonstrating ADOPT: Adaptively Optimizing Attribute Orders for Worst-Case Optimal Joins via Reinforcement Learning.....	4094
<i>Junxiong Wang, Mitchell Gray, Immanuel Trummer, Ahmet Kara, Dan Olteanu</i>	
Demonstrating GPT-DB: Generating Query-Specific and Customizable Code for SQL Processing with GPT-4.....	4098
<i>Immanuel Trummer</i>	
SHEVA: A Visual Analytics System for Statistical Hypothesis Exploration.....	4102
<i>Vicente Nejar De Almeida, Eduardo Ribeiro, Nassim Bouarour, João Luiz Dihl Comba, Sihem Amer-Yahia</i>	
PikePlace: Generating Intelligence for Marketplace Datasets.....	4106
<i>Shi Qiao, Alekh Jindal</i>	



A Learned Query Rewrite System .....	4110
<i>Xuanhe Zhou, Guoliang Li, Jianming Wu, Jiesi Liu, Zhaoyan Sun, Xinning Zhang</i>	

### **Panel**

Will LLMs reshape, supercharge, or kill data science? .....	4114
<i>Alon Halevy, Yejin Choi, Avriella Floratou, Michael J. Franklin, Natasha Noy, Haixun Wang</i>	

### **Endowment Awards**

Towards Auto-Generated Data Systems .....	4116
<i>Alvin Cheung, Maaz Bin Safeer Ahmad, Brandon Haynes, Chanwut Kittivorawong, Shadaj Laddad, Xiaoxuan Liu, Chenglong Wang, Cong Yan</i>	

Generations of Knowledge Graphs: The Crazy Ideas and the Business Impact .....	4130
<i>Xin Luna Dong</i>	

The Story of GraphLab – From Scaling Machine Learning to Shaping Graph Systems Research .....	4138
<i>Joseph E. Gonzalez, Yucheng Low</i>	

### **Keynotes**

Common Sense: the Dark Matter of Language and Intelligence .....	4139
<i>Yejin Choi</i>	

Modernization of Databases in the Cloud Era: Building Databases that Run Like Legos .....	4140
<i>Feifei Li</i>	

On the Cusp: Computing Thrills and Perils and Professional Awakening .....	4152
<i>Natasa Milic-Frayling</i>	

## TABLE OF CONTENTS

### Research Papers

Doquet: Differentially Oblivious Range and Join Queries with Private Data Structures .....	4160
<i>Lina Qiu, Georgios Kellaris, Nikos Mamoulis, Kobbi Nissim, George Kollios</i>	
AMNES: Accelerating the computation of data correlation using FPGAs.....	4174
<i>Monica Chiosa, Thomas B. Preußner, Michaela Blott, Gustavo Alonso</i>	
VOCALExplore: Pay-as-You-Go Video Data Exploration and Model Building .....	4188
<i>Maureen Daum, Enhao Zhang, Dong He, Stephen Mussmann, Brandon Haynes, Ranjay Krishna, Magdalena Balazinska</i>	
Flexible Resource Allocation for Relational Database-as-a-Service .....	4202
<i>Pankaj Arora, Surajit Chaudhuri, Sudipto Das, Junfeng Dong, Cyril George, Ajay Kalhan, Arnd Christian König, Willis Lang, Changsong Li, Feng Li, Jiaqi Liu, Lukas M. Maas, Akshay Mata, Ishai Menache, Justin Moeller, Vivek Narasayya, Matthaios Olma, Morgan Oslake, Elnaz Rezai, Yi Shan, Manoj Syamala, Shize Xu, Vasileios Zois</i>	
ShadowAQP: Efficient Approximate Group-by and Join Query via Attribute-oriented Sample Size Allocation and Data Generation.....	4216
<i>Rong Gu, Han Li, Haipeng Dai, Wenjie Huang, Jie Xue, Meng Li, Jiaqi Zheng, Haoran Cai, Yihua Huang, Guihai Chen</i>	
Optimizing Data Pipelines for Machine Learning in Feature Stores .....	4230
<i>Rui Liu, Kwanghyun Park, Fotis Psallidas, Xiaoyong Zhu, Jinghui Mo, Rathijit Sen, Matteo Interlandi, Konstantinos Karanasos, Yuanyuan Tian, Jesús Camacho-Rodríguez</i>	
SparqLog: A System for Efficient Evaluation of SPARQL 1.1 Queries via Datalog .....	4240
<i>Renzo Angles, Georg Gottlob, Aleksandar Pavlović, Reinhard Pichler, Emanuel Sallinger</i>	
Solver-In-The-Loop Cluster Resource Management for Database-as-a-Service .....	4254
<i>Arnd Christian König, Yi Shan, Karan Newatia, Luke Marshall, Vivek Narasayya</i>	
RTIndex: Exploiting Hardware-Accelerated GPU Raytracing for Database Indexing .....	4268
<i>Justus Henneberg, Felix Schuhknecht</i>	
ContTune: Continuous Tuning by Conservative Bayesian Optimization for Distributed Stream Data Processing Systems .....	4282
<i>Jinqing Lian, Xinyi Zhang, Yingxia Shao, Zenglin Pu, Qingfeng Xiang, Yawen Li, Bin Cui</i>	
Single Update Sketch with Variable Counter Structure .....	4296
<i>Dimitrios Melissourgios, Haibo Wang, Shigang Chen, Chaoyi Ma, Shiping Chen</i>	
Can Large Language Models Predict Data Correlations from Column Names?.....	4310
<i>Immanuel Trummer</i>	
GraphOS: Towards Oblivious Graph Processing .....	4324
<i>Javad Ghareh Chamani, Ioannis Demertzis, Dimitrios Papadopoulos, Charalampos Papamanthou, Rasool Jalili</i>	

Catalyst: Optimizing Cache Management for Large In-memory Key-value Systems..... 4339  
*Kefei Wang, Feng Chen*