

50th International Conference on Very Large Data Bases (VLDB 2024)

Proceedings of the VLDB Endowment Volume 17

Guangzhou, China
26-30 August 2024

Part 1 of 7

ISBN: 979-8-3313-0949-7

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© (2024) 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.

Printed with permission by Curran Associates, Inc. (2025)

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.

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
Web: www.proceedings.com

TABLE OF CONTENTS

Research Papers

DecLog: Decentralized Logging in Non-Volatile Memory for Time Series Database Systems	1
<i>Bolong Zheng, Yongyong Gao, Jingyi Wan, Lingsen Yan, Long Hu, Bo Liu, Yunjun Gao, Xiaofang Zhou, Christian S. Jensen</i>	
Efficient Dynamic Weighted Set Sampling and Its Extension	15
<i>Fangyuan Zhang, Mengxu Jiang, Sibao Wang</i>	
ZIP: Lazy Imputation during Query Processing	28
<i>Yiming Lin, Sharad Mehrotra</i>	
FedGTA: Topology-aware Averaging for Federated Graph Learning.....	41
<i>Xunkai Li, Zhengyu Wu, Wentao Zhang, Yinlin Zhu, Ronghua Li, Guoren Wang</i>	
Host Profit Maximization: Leveraging Performance Incentives and User Flexibility	51
<i>Xueqin Chang, Xiangyu Ke, Lu Chen, Congcong Ge, Ziheng Wei, Yunjun Gao</i>	
DP-PQD: Privately Detecting Per-Query Gaps In Synthetic Data Generated By Black-Box Mechanisms	65
<i>Shweta Patwa, Danyu Sun, Amir Gilad, Ashwin Machanavajjhala, Sudeepa Roy</i>	

TABLE OF CONTENTS

Research Papers

Cryptographically Secure Private Record Linkage Using Locality-Sensitive Hashing.....	79
<i>Ruidi Wei, Florian Kerschbaum</i>	
Language Models Enable Simple Systems for Generating Structured Views of Heterogeneous Data Lakes	92
<i>Simran Arora, Brandon Yang, Sabri Eyuboglu, Avaniika Narayan, Andrew Hojel, Immanuel Trummer, Christopher Re</i>	
Query Refinement for Diversity Constraint Satisfaction.....	106
<i>Jinyang Li, Yuval Moskovitch, Julia Stoyanovich, H. V. Jagadish</i>	
ElasticNotebook: Enabling Live Migration for Computational Notebooks	119
<i>Zhaoheng Li, Pranav Gor, Rahul Prabhu, Hui Yu, Yuzhou Mao, Yongjoo Park</i>	
Breathing New Life into An Old Tree: Resolving Logging Dilemma of $B^+{-}$ tree on Modern Computational Storage Drives.....	134
<i>Kecheng Huang, Zhaoyan Shen, Zili Shao, Tong Zhang, Feng Chen</i>	
An Empirical Evaluation of Columnar Storage Formats	148
<i>Xinyu Zeng, Yulong Hui, Jiahong Shen, Andrew Pavlo, Wes McKinney, Huanchen Zhang</i>	
Everest: GPU-Accelerated System For Mining Temporal Motifs	162
<i>Yichao Yuan, Haojie Ye, Sanketh Vedula, Wynn M Kaza, Nishil Talati</i>	
Billion-Scale Bipartite Graph Embedding: A Global-Local Induced Approach.....	175
<i>Xueyi Wu, Yuanyuan Xu, Wenjie Zhang, Ying Zhang</i>	
Utility-aware Payment Channel Network Rebalance	184
<i>Wangze Ni, Pengze Chen, Lei Chen, Peng Cheng, Chen Zhang, Xuemin Lin</i>	
ALECE: An Attention-based Learned Cardinality Estimator for SPJ Queries on Dynamic Workloads.....	197
<i>Pengfei Li, Wenqing Wei, Rong Zhu, Bolin Ding, Jingren Zhou, Hua Lu</i>	
Flash-LLM: Enabling Low-Cost and Highly-Efficient Large Generative Model Inference With Unstructured Sparsity	211
<i>Haojun Xia, Zhen Zheng, Yuchao Li, Donglin Zhuang, Zhongzhu Zhou, Xiafei Qiu, Yong Li, Wei Lin, Shuaiwen Leon Song</i>	
Confidential Consortium Framework: Secure Multiparty Applications with Confidentiality, Integrity, and High Availability	225
<i>Heidi Howard, Fritz Alder, Edward Ashton, Amaury Chamayou, Sylvan Clebsch, Manuel Costa, Antoine Delignat-Lavaud, Cedric Fournet, Andrew Jeffery, Matthew Kerner, Fotios Kounelis, Markus A. Kuppe, Julien Maffre, Mark Russinovich, Christoph M. Wintersteiger</i>	
VeLP: Vehicle Loading Plan Learning from Human Behavior in Nationwide Logistics System	241
<i>Sijing Duan, Feng Lyu, Xin Zhu, Yi Ding, Haotian Wang, Desheng Zhang, Xue Liu, Yaoyue Zhang, Ju Ren</i>	

Relational Query Synthesis $\bowtie$ Decision Tree Learning.....	250
<i>Aaditya Naik, Aalok Thakkar, Adam Stein, Rajeev Alur, Mayur Naik</i>	

TABLE OF CONTENTS

Research Papers

RAGraph: A Region-Aware Framework for Geo-Distributed Graph Processing	264
<i>Feng Yao, Qian Tao, Wenyuan Yu, Yanfeng Zhang, Shufeng Gong, Qiang Wang, Ge Yu, Jingren Zhou</i>	
SmartLite: A DBMS-based Serving System for DNN Inference in Resource-constrained Environments	278
<i>Qiuru Lin, Sai Wu, Junbo Zhao, Jian Dai, Meng Shi, Gang Chen, Feifei Li</i>	
Blocker and Matcher Can Mutually Benefit: A Co-Learning Framework for Low-Resource Entity Resolution	292
<i>Shiwen Wu, Qiyu Wu, Honghua Dong, Wen Hua, Xiaofang Zhou</i>	
TSGBench: Time Series Generation Benchmark.....	305
<i>Yihao Ang, Qiang Huang, Yifan Bao, Anthony K. H. Tung, Zhiyong Huang</i>	
OmniSketch: Efficient Multi-Dimensional High-Velocity Stream Analytics with Arbitrary Predicates	319
<i>Wieger R. Punter, Odysseas Papapetrou, Minos Garofalakis</i>	
Maximum Balanced (k, ϵ) -Bitruss Detection in Signed Bipartite Graph	332
<i>Kai Hiu Chung, Alexander Zhou, Yue Wang, Lei Chen</i>	
Missing Value Imputation for Multi-attribute Sensor Data Streams via Message Propagation	345
<i>Xiao Li, Huan Li, Hua Lu, Christian S. Jensen, Varun Pandey, Volker Markl</i>	
ImDiffusion: Imputed Diffusion Models for Multivariate Time Series Anomaly Detection	359
<i>Yuhang Chen, Chaoyun Zhang, Minghua Ma, Yudong Liu, Ruomeng Ding, Bowen Li, Shilin He, Saravan Rajmohan, Qingwei Lin, Dongmei Zhang</i>	
Confidence Intervals for Private Query Processing	373
<i>Dajun Sun, Wei Dong, Ke Yi</i>	
A Shapelet-based Framework for Unsupervised Multivariate Time Series Representation Learning	386
<i>Zhiyu Liang, Jianfeng Zhang, Chen Liang, Hongzhi Wang, Zheng Liang, Lujia Pan</i>	
Fast and Space-Efficient Parallel Algorithms for Influence Maximization.....	400
<i>Letong Wang, Xiangyun Ding, Yan Gu, Yihan Sun</i>	
TERI: An Effective Framework for Trajectory Recovery with Irregular Time Intervals	414
<i>Yile Chen, Gao Cong, Cuauhtemoc Anda</i>	
Demystifying Graph Sparsification Algorithms in Graph Properties Preservation	427
<i>Yuhan Chen, Haojie Ye, Sanketh Vedula, Alex Bronstein, Ronald Dreslinski, Trevor Mudge, Nishil Talati</i>	
GPU Database Systems Characterization and Optimization.....	441
<i>Jiashen Cao, Rathijit Sen, Matteo Interlandi, Joy Arulraj, Hyesoon Kim</i>	

NeutronStream: A Dynamic GNN Training Framework with Sliding Window for Graph Streams	455
<i>Chaoyi Chen, Dechao Gao, Yanfeng Zhang, Qiange Wang, Zhenbo Fu, Xuechang Zhang, Junhua Zhu, Yu Gu, Ge Yu</i>	
Caerus: Low-Latency Distributed Transactions for Geo-Replicated Systems	469
<i>Joshua Hildred, Michael Abebe, Khuzaima Daudjee</i>	
An Experimental Evaluation of Anomaly Detection in Time Series	483
<i>Aoqian Zhang, Shuqing Deng, Dongping Cui, Ye Yuan, Guoren Wang</i>	
FormaT5: Abstention and Examples for Conditional Table Formatting with Natural Language	497
<i>Mukul Singh, José Cambrono, Sumit Gulwani, Vu Le, Carina Negreanu, Elnaz Nouri, Mohammad Raza, Gust Verbruggen</i>	
Quantum-Inspired Digital Annealing for Join Ordering	511
<i>Manuel Schönberger, Immanuel Trummer, Wolfgang Mauerer</i>	
KAMEL: A Scalable BERT-based System for Trajectory Imputation.....	525
<i>Mashaal Musleh, Mohamed F. Mokbel</i>	
An Efficient Transfer Learning Based Configuration Adviser for Database Tuning	539
<i>Xinyi Zhang, Hong Wu, Yang Li, Zhengju Tang, Jian Tan, Feifei Li, Bin Cui</i>	
ADF & TransApp: A Transformer-Based Framework for Appliance Detection Using Smart Meter Consumption Series.....	553
<i>Adrien Petralia, Philippe Charpentier, Themis Palpanas</i>	
RALF: Accuracy-Aware Scheduling for Feature Store Maintenance.....	563
<i>Sarah Wooders, Xiangxi Mo, Amit Narang, Kevin Lin, Ion Stoica, Joseph M. Hellerstein, Natacha Crooks, Joseph E. Gonzalez</i>	
The Art of Latency Hiding in Modern Database Engines	577
<i>Kaisong Huang, Tianzheng Wang, Qingqing Zhou, Qingzhong Meng</i>	
MOSER: Scalable Network Motif Discovery using Serial Test	591
<i>Mohammad Matin Najafi, Chenhao Ma, Xiaodong Li, Reynold Cheng, Laks V.S. Lakshmanan</i>	
Co-movement Pattern Mining from Videos	604
<i>Dongxiang Zhang, Teng Ma, Junnan Hu, Yijun Bei, Kian-Lee Tan, Gang Chen</i>	

TABLE OF CONTENTS

Research Papers

Efficient and Accurate SimRank-based Similarity Joins: Experiments, Analysis, and Improvement.....	617
<i>Qian Ge, Yu Liu, Yinghao Zhao, Yuetian Sun, Lei Zou, Yuxing Chen, Anqun Pan</i>	
Expanding Reverse Nearest Neighbors	630
<i>Wentao Li, Maolin Cai, Min Gao, Dong Wen, Lu Qin, Wei Wang</i>	
Errata for "SpaceSaving\$: An Optimal Algorithm for Frequency Estimation and Frequent Items in the Bounded-Deletion Model"	643
<i>Fuheng Zhao, Divyakant Agrawal, Amr El Abbadi, Ahmed Metwally, Claire Mathieu, Michel De Rougemont</i>	
Cache-Efficient Top-k Aggregation over High Cardinality Large Datasets	644
<i>Tarique Siddiqui, Vivek Narasayya, Marius Dumitru, Surajit Chaudhuri</i>	
Efficient Temporal Butterfly Counting and Enumeration on Temporal Bipartite Graphs.....	657
<i>Xinwei Cai, Xiangyu Ke, Kai Wang, Lu Chen, Tianming Zhang, Qing Liu, Yunjun Gao</i>	
TVM: A Tile-based Video Management Framework	671
<i>Tianxiong Zhong, Zhiwei Zhang, Guo Lu, Ye Yuan, Yu-Ping Wang, Guoren Wang</i>	
ScienceBenchmark: A Complex Real-World Benchmark for Evaluating Natural Language to SQL Systems	685
<i>Yi Zhang, Jan Deriu, George Katsogiannis-Meimarakis, Catherine Kosten, Georgia Koutrika, Kurt Stockinger</i>	
Densest Multipartite Subgraph Search in Heterogeneous Information Networks	699
<i>Lu Chen, Chengfei Liu, Rui Zhou, Kewen Liao, Jiajie Xu, Jianxin Li</i>	
Saturn: An Optimized Data System for Multi-Large-Model Deep Learning Workloads.....	712
<i>Kabir Nagrecha, Arun Kumar</i>	
BonsaiKV: Towards Fast, Scalable, and Persistent Key-Value Stores with Tiered, Heterogeneous Memory System	726
<i>Miao Cai, Junru Shen, Yifan Yuan, Zhihao Qu, Baoliu Ye</i>	
Sample-Efficient Cardinality Estimation Using Geometric Deep Learning.....	740
<i>Silvan Reiner, Michael Grossniklaus</i>	
Multiple Time Series Forecasting with Dynamic Graph Modeling	753
<i>Kai Zhao, Chenjuan Guo, Yunyao Cheng, Peng Han, Miao Zhang, Bin Yang</i>	
Weakly Guided Adaptation for Robust Time Series Forecasting	766
<i>Yunyao Cheng, Peng Chen, Chenjuan Guo, Kai Zhao, Qingsong Wen, Bin Yang, Christian S. Jensen</i>	
Algorithmic Complexity Attacks on Dynamic Learned Indexes.....	780
<i>Rui Yang, Evgenios M. Kornaropoulos, Yue Cheng</i>	

METER: A Dynamic Concept Adaptation Framework for Online Anomaly Detection	794
<i>Jiaqi Zhu, Shaofeng Cai, Fang Deng, Beng Chin Ooi, Wenqiao Zhang</i>	
Experimental Analysis of Large-scale Learnable Vector Storage Compression	808
<i>Hailin Zhang, Penghao Zhao, Xupeng Miao, Yingxia Shao, Zirui Liu, Tong Yang, Bin Cui</i>	
A Comparative Study and Component Analysis of Query Plan Representation Techniques in ML4DB Studies	823
<i>Yue Zhao, Zhaodonghui Li, Gao Cong</i>	
Testing Graph Database Systems via Graph-Aware Metamorphic Relations	836
<i>Zeyang Zhuang, Penghui Li, Pingchuan Ma, Wei Meng, Shuai Wang</i>	
Observatory: Characterizing Embeddings of Relational Tables	849
<i>Tianji Cong, Madelon Hulsebos, Zhenjie Sun, Paul Groth, H. V. Jagadish</i>	
FusionFlow: Accelerating Data Preparation for Machine Learning with Hybrid CPU-GPU Processing	863
<i>Taeyoon Kim, ChanHo Park, Mansur Mukimbekov, Heelim Hong, Minseok Kim, Ze Jin, Changdae Kim, Ji-Yong Shin, Myeongjae Jeon</i>	
BOSS - An Architecture for Database Kernel Composition	877
<i>Hubert Mohr-Daurat, Xuan Sun, Holger Pirk</i>	
CoroGraph: Bridging Cache Efficiency and Work Efficiency for Graph Algorithm Execution	891
<i>Xiangyu Zhi, Xiao Yan, Bo Tang, Ziyao Yin, Yanchao Zhu, Minqi Zhou</i>	
Mammoths Are Slow: The Overlooked Transactions of Graph Data	904
<i>Audrey Cheng, Jack Waudby, Hugo Firth, Natacha Crooks, Ion Stoica</i>	
VeriDKG: A Verifiable SPARQL Query Engine for Decentralized Knowledge Graphs	912
<i>Enyuan Zhou, Song Guo, Zicong Hong, Christian S. Jensen, Yang Xiao, Dalin Zhang, Jinwen Liang, Qingqi Pei</i>	

TABLE OF CONTENTS

Research Papers

Eraser: Eliminating Performance Regression on Learned Query Optimizer	926
<i>Lianggui Weng, Rong Zhu, Di Wu, Bolin Ding, Bolong Zheng, Jingren Zhou</i>	
HyBench: A New Benchmark for HTAP Databases	939
<i>Chao Zhang, Guoliang Li, Tao Lv</i>	
Falcon: Fair Active Learning using Multi-armed Bandits	952
<i>Ki Hyun Tae, Hantian Zhang, Jaeyoung Park, Kexin Rong, Steven Euijong Whang</i>	
A Blockchain System for Clustered Federated Learning with Peer-to-Peer Knowledge Transfer	966
<i>Honghu Wu, Xiangrong Zhu, Wei Hu</i>	
PilotScope: Steering Databases with Machine Learning Drivers	980
<i>Rong Zhu, Lianggui Weng, Wenqing Wei, Di Wu, Jiazhen Peng, Yifan Wang, Bolin Ding, Defu Lian, Bolong Zheng, Jingren Zhou</i>	
Timestamp as a Service, not an Oracle	994
<i>Yishuai Li, Yunfeng Zhu, Chao Shi, Guanhua Zhang, Jianzhong Wang, Xiaolu Zhang</i>	
Data-Driven Insight Synthesis for Multi-Dimensional Data	1007
<i>Junjie Xing, Xinyu Wang, H. V. Jagadish</i>	
Database Native Model Selection: Harnessing Deep Neural Networks in Database Systems	1020
<i>Naili Xing, Shaofeng Cai, Gang Chen, Zhaojing Luo, Beng Chin Ooi, Jian Pei</i>	
Querying Structural Diversity in Streaming Graphs	1034
<i>Kaiyu Chen, Dong Wen, Wenjie Zhang, Ying Zhang, Xiaoyang Wang, Xuemin Lin</i>	
LM-SRPQ: Efficiently Answering Regular Path Query in Streaming Graphs	1047
<i>Xiangyang Gou, Xinyi Ye, Lei Zou, Jeffrey Xu Yu</i>	
ETC: Efficient Training of Temporal Graph Neural Networks over Large-scale Dynamic Graphs	1060
<i>Shihong Gao, Yiming Li, Yanyan Shen, Yingxia Shao, Lei Chen</i>	
Towards Full Stack Adaptivity in Permissioned Blockchains	1073
<i>Chenyuan Wu, Mohammad Javad Amiri, Haoyun Qin, Bhavana Mehta, Ryan Marcus, Boon Thau Loo</i>	
BigST: Linear Complexity Spatio-Temporal Graph Neural Network for Traffic Forecasting on Large-Scale Road Networks	1081
<i>Jindong Han, Weijia Zhang, Hao Liu, Tao Tao, Naiqiang Tan, Hui Xiong</i>	
SepHash: A Write-Optimized Hash Index On Disaggregated Memory via Separate Segment Structure... ..	1091
<i>Xinhao Min, Kai Lu, Pengyu Liu, Jiguang Wan, Changsheng Xie, Daohui Wang, Ting Yao, Huatao Wu</i>	

XGNN: Boosting Multi-GPU GNN Training via Global GNN Memory Store	1105
<i>Dahai Tang, Jiali Wang, Rong Chen, Lei Wang, Wenyuan Yu, Jingren Zhou, Kenli Li</i>	
Communication Efficient and Provable Federated Unlearning	1119
<i>Youming Tao, Cheng-Long Wang, Miao Pan, Dongxiao Yu, Xiuzhen Cheng, Di Wang</i>	
Text-to-SQL Empowered by Large Language Models: A Benchmark Evaluation	1132
<i>Dawei Gao, Haibin Wang, Yaliang Li, Xiuyu Sun, Yichen Qian, Bolin Ding, Jingren Zhou</i>	
Scaling Package Queries to a Billion Tuples via Hierarchical Partitioning and Customized Optimization	1146
<i>Anh L. Mai, Pengyu Wang, Azza Abouzied, Matteo Brucato, Peter J. Haas, Alexandra Meliou</i>	

TABLE OF CONTENTS

Research Papers

MisDetect: Iterative Mislabel Detection using Early Loss	1159
<i>Yuhao Deng, Chengliang Chai, Lei Cao, Nan Tang, Jiayi Wang, Ju Fan, Ye Yuan, Guoren Wang</i>	
Capturing More Associations by Referencing External Graphs	1173
<i>Wenfei Fan, Muyang Liu, Shuhao Liu, Chao Tian</i>	
QTCS: Efficient Query-Centered Temporal Community Search.....	1187
<i>Longlong Lin, Pingpeng Yuan, Rong-Hua Li, Chunxue Zhu, Hongchao Qin, Hai Jin, Tao Jia</i>	
DPSUR: Accelerating Differentially Private Stochastic Gradient Descent Using Selective Update and Release	1200
<i>Jie Fu, Qingqing Ye, Haibo Hu, Zhili Chen, Lulu Wang, Kuncan Wang, Xun Ran</i>	
How Can We Train Deep Learning Models Across Clouds and Continents? An Experimental Study	1214
<i>Alexander Erben, Ruben Mayer, Hans-Arno Jacobsen</i>	
Accelerating Sampling and Aggregation Operations in GNN Frameworks with GPU Initiated Direct Storage Accesses	1227
<i>Jeongmin Brian Park, Vikram Sharma Mailthody, Zaid Qureshi, Wen-Mei Hwu</i>	
Comprehensive Evaluation of GNN Training Systems: A Data Management Perspective	1241
<i>Hao Yuan, Yajiong Liu, Yanfeng Zhang, Xin Ai, Qiang Wang, Chaoyi Chen, Yu Gu, Ge Yu</i>	
LION: Fast and High-Resolution Network Kernel Density Visualization	1255
<i>Tsz Nam Chan, Rui Zang, Bojian Zhu, Leong Hou U, Dingming Wu, Jianliang Xu</i>	
Performance-Based Pricing of Federated Learning via Auction	1269
<i>Zitao Li, Bolin Ding, Liuyi Yao, Yaliang Li, Xiaokui Xiao, Jingren Zhou</i>	
OE Bench: Investigating Open Environment Challenges in Real-World Relational Data Streams	1283
<i>Yiqun Diao, Yutong Yang, Qinbin Li, Bingsheng He, Mian Lu</i>	
Influence Maximization via Vertex Countering.....	1297
<i>Jiadong Xie, Zehua Chen, Deming Chu, Fan Zhang, Xuemin Lin, Zhihong Tian</i>	
Optimizing Data Acquisition to Enhance Machine Learning Performance.....	1310
<i>Tingting Wang, Shixun Huang, Zhifeng Bao, J. Shane Culpepper, Volkan Dedeoglu, Reza Arablouei</i>	
Minimum Strongly Connected Subgraph Collection in Dynamic Graphs	1324
<i>Xin Chen, Jieming Shi, You Peng, Wenqing Lin, Sibao Wang, Wenjie Zhang</i>	
FusionQuery: On-demand Fusion Queries over Multi-source Heterogeneous Data	1337
<i>Junhao Zhu, Yuren Mao, Lu Chen, Congcong Ge, Ziheng Wei, Yunjun Gao</i>	

POLAR: Adaptive and Non-invasive Join Order Selection via Plans of Least Resistance.....	1350
<i>David Justen, Daniel Ritter, Campbell Fraser, Andrew Lamb, Nga Tran, Allison Lee, Thomas Bodner, Mhd Yamen Haddad, Steffen Zeuch, Volker Markl, Matthias Boehm</i>	
DAHA: Accelerating GNN Training with Data and Hardware Aware Execution Planning.....	1364
<i>Zhiyuan Li, Xun Jian, Yue Wang, Yingxia Shao, Lei Chen</i>	
FluidKV: Seamlessly Bridging the Gap between Indexing Performance and Memory-Footprint on Ultra-Fast Storage.....	1377
<i>Ziyi Lu, Qiang Cao, Hong Jiang, Yuxing Chen, Jie Yao, Anqun Pan</i>	
How do Categorical Duplicates Affect ML? A New Benchmark and Empirical Analyses	1391
<i>Vraj Shah, Thomas J Parashos, Arun Kumar</i>	
CGgraph: An Ultra-fast Graph Processing System on Modern Commodity CPU-GPU Co-processor	1405
<i>Pengjie Cui, Haotian Liu, Bo Tang, Ye Yuan</i>	
FCBench: Cross-Domain Benchmarking of Lossless Compression for Floating-point Data	1418
<i>Xinyu Chen, Jiannan Tian, Ian Beaver, Cynthia Freeman, Yan Yan, Jianguo Wang, Dingwen Tao</i>	
PairwiseHist: Fast, Accurate, and Space-Efficient Approximate Query Processing with Data Compression	1432
<i>Aaron Hurst, Daniel E Lucani, Qi Zhang</i>	
MetaStore: Analyzing Deep Learning Meta-Data at Scale.....	1446
<i>Huayi Zhang, Binwei Yan, Lei Cao, Samuel Madden, Elke Rundensteiner</i>	
RTScan: Efficient Scan with Ray Tracing Cores.....	1460
<i>Yangming Lv, Kai Zhang, Ziming Wang, Xiaodong Zhang, Rubao Lee, Zhenying He, Yinan Jing, X. Sean Wang</i>	
FreshGNN: Reducing Memory Access via Stable Historical Embeddings for Graph Neural Network Training	1473
<i>Kezhao Huang, Haitian Jiang, Minjie Wang, Guangxuan Xiao, David Wipf, Xiang Song, Quan Gan, Zengfeng Huang, Jidong Zhai, Zheng Zhang</i>	
Sorting on Byte-Addressable Storage: The Resurgence of Tree Structure	1487
<i>Ying Zheng, Kian-Lee Tan</i>	
Efficient Placement of Decomposable Aggregation Functions for Stream Processing over Large Geo-Distributed Topologies.....	1501
<i>Xenofon Chatziliadis, Eleni Tzirita Zacharatou, Alphan Eracar, Steffen Zeuch, Volker Markl</i>	
AeonG: An Efficient Built-in Temporal Support in Graph Databases	1515
<i>Jiamin Hou, Zhanhao Zhao, Zhouyu Wang, Wei Lu, Guodong Jin, Dong Wen, Xiaoyong Du</i>	

TABLE OF CONTENTS

Research Papers

Refactoring Index Tuning Process with Benefit Estimation.....	1528
<i>Tao Yu, Zhaonian Zou, Weihua Sun, Yu Yan</i>	
LightDiC: A Simple yet Effective Approach for Large-scale Digraph Representation Learning	1542
<i>Xunkai Li, Meihao Liao, Zhengyu Wu, Daohan Su, Wentao Zhang, Ronghua Li, Guoren Wang</i>	
Efficient Differential Dependency Discovery	1552
<i>Shulei Kuang, Honghui Yang, Zijing Tan, Shuai Ma</i>	
Is Your Learned Query Optimizer Behaving As You Expect? A Machine Learning Perspective	1565
<i>Claude Lehmann, Pavel Sulimov, Kurt Stockinger</i>	
Mixed Covers of Keys and Functional Dependencies for Maintaining the Integrity of Data under Updates	1578
<i>Zhuoxing Zhang, Sebastian Link</i>	
Outlier Summarization via Human Interpretable Rules	1591
<i>Yuhao Deng, Yu Wang, Lei Cao, Lianpeng Qiao, Yuping Wang, Jingzhe Xu, Yizhou Yan, Samuel Madden</i>	
Nuhuo: An Effective Estimation Model for Traffic Speed Histogram Imputation on A Road Network.....	1605
<i>Haitao Yuan, Gao Cong, Guoliang Li</i>	
Intelligent Pooling: Proactive Resource Provisioning in Large-scale Cloud Service	1618
<i>Deepak Ravikumar, Alex Yeo, Yiwen Zhu, Aditya Lakra, Harsha Nagulapalli, Santhosh Ravindran, Steve Suh, Niharika Dutta, Andrew Fogarty, Yoonjae Park, Sumeet Khushalani, Arijit Tarafdar, Kunal Parekh, Subru Krishnan</i>	
Efficient Exact Subgraph Matching via GNN-based Path Dominance Embedding.....	1628
<i>Yutong Ye, Xiang Lian, Mingsong Chen</i>	
Leveraging Dynamic and Heterogeneous Workload Knowledge to Boost the Performance of Index Advisors	1642
<i>Zijia Wang, Haoran Liu, Chen Lin, Zhifeng Bao, Guoliang Li, Tianqing Wang</i>	
UltraLogLog: A Practical and More Space-Efficient Alternative to HyperLogLog for Approximate Distinct Counting.....	1655
<i>Otmar Ertl</i>	
Real-time Insertion Operator for Shared Mobility on Time-Dependent Road Networks	1669
<i>Zengyang Gong, Yuxiang Zeng, Lei Chen</i>	
X-TED: Massive Parallelization of Tree Edit Distance	1683
<i>Dayi Fan, Rubao Lee, Xiaodong Zhang</i>	

Cardinality Estimation of Subgraph Matching: A Filtering-Sampling Approach.....	1697
<i>Wonseok Shin, Siwoo Song, Kunsoo Park, Wook-Shin Han</i>	
Efficient Regular Simple Path Queries under Transitive Restricted Expressions	1710
<i>Qi Liang, Dian Ouyang, Fan Zhang, Jianye Yang, Xuemin Lin, Zhihong Tian</i>	
A Multi-Scale Decomposition MLP-Mixer for Time Series Analysis.....	1723
<i>Shuhan Zhong, Sizhe Song, Weipeng Zhuo, Guanyao Li, Yang Liu, S.-H. Gary Chan</i>	
P-Shapley: Shapley Values on Probabilistic Classifiers	1737
<i>Haocheng Xia, Xiang Li, Junyuan Pang, Jinfei Liu, Kui Ren, Li Xiong</i>	
Optimizing Video Selection LIMIT Queries With Commonsense Knowledge.....	1751
<i>Wenjia He, Ibrahim Sabek, Yuze Lou, Michael Cafarella</i>	
ZeroEA: A Zero-Training Entity Alignment Framework via Pre-Trained Language Model	1765
<i>Nan Huo, Reynold Cheng, Ben Kao, Wentao Ning, Nur Al Hasan Haldar, Xiaodong Li, Jinyang Li, Mohammad Matin Najafi, Tian Li, Ge Qu</i>	
Extending Graph Rules with Oracles.....	1775
<i>Xueli Liu, Bowen Dong, Wenzhi Fu, Nannan Wu, Xin Wang, Wenjun Wang</i>	

TABLE OF CONTENTS

Research Papers

FlowWalker: A Memory-efficient and High-performance GPU-based Dynamic Graph Random Walk Framework.....	1788
<i>Junyi Mei, Shixuan Sun, Chao Li, Cheng Xu, Cheng Chen, Yibo Liu, Jing Wang, Cheng Zhao, Xiaofeng Hou, Minyi Guo, Bingsheng He, Xiaoliang Cong</i>	
Accelerating String-key Learned Index Structures via Memoization-based Incremental Training	1802
<i>Minsu Kim, Jinwoo Hwang, Guseul Heo, Seiyen Cho, Divya Mahajan, Jongse Park</i>	
Truss-based Community Search over Streaming Directed Graphs	1816
<i>Xuankun Liao, Qing Liu, Xin Huang, Jianliang Xu</i>	
InferDB: In-Database Machine Learning Inference Using Indexes.....	1830
<i>Ricardo Salazar-Díaz, Boris Glavic, Tilmann Rabl</i>	
AAA: an Adaptive Mechanism for Locally Differential Private Mean Estimation	1843
<i>Fei Wei, Ergute Bao, Xiaokui Xiao, Yin Yang, Bolin Ding</i>	
Accelerating Merkle Patricia Trie with GPU	1856
<i>Yangshen Deng, Muxi Yan, Bo Tang</i>	
Privacy Amplification via Shuffling: Unified, Simplified, and Tightened	1870
<i>Shaowei Wang, Yun Peng, Jin Li, Zikai Wen, Zhipeng Li, Shiyu Yu, Di Wang, Wei Yang</i>	
Detecting Metadata-Related Logic Bugs in Database Systems via Raw Database Construction.....	1884
<i>Jiansen Song, Wensheng Dou, Yu Gao, Ziyu Cui, Yingying Zheng, Dong Wang, Wei Wang, Jun Wei, Tao Huang</i>	
From Zero to Hero: Detecting Leaked Data through Synthetic Data Injection and Model Querying.....	1898
<i>Biao Wu, Qiang Huang, Anthony K. H. Tung</i>	
Oasis: An Optimal Disjoint Segmented Learned Range Filter	1911
<i>Guanduo Chen, Zhenying He, Meng Li, Siqiang Luo</i>	
LakeBench: A Benchmark for Discovering Joinable and Unionable Tables in Data Lakes	1925
<i>Yuhao Deng, Chengliang Chai, Lei Cao, Qin Yuan, Siyuan Chen, Yanrui Yu, Zhaoze Sun, Junyi Wang, Jiajun Li, Ziqi Cao, Kaisen Jin, Chi Zhang, Yuqing Jiang, Yuanfang Zhang, Yuping Wang, Ye Yuan, Guoren Wang, Nan Tang</i>	
GPTuner: A Manual-Reading Database Tuning System via GPT-Guided Bayesian Optimization.....	1939
<i>Jiale Lao, Yibo Wang, Yufei Li, Jianping Wang, Yunjia Zhang, Zhiyuan Cheng, Wanghu Chen, Mingjie Tang, Jianguo Wang</i>	
Raising the ClaSS of Streaming Time Series Segmentation.....	1953
<i>Arik Ermshaus, Patrick Schäfer, Ulf Leser</i>	

Fast Local Subgraph Counting	1967
<i>Qiyan Li, Jeffrey Xu Yu</i>	
ReAcTable: Enhancing ReAct for Table Question Answering.....	1981
<i>Yunjia Zhang, Jordan Henkel, Avrilia Floratou, Joyce Cahoon, Shaleen Deep, Jignesh M. Patel</i>	
NeutronOrch: Rethinking Sample-based GNN Training under CPU-GPU Heterogeneous Environments	1995
<i>Xin Ai, Qiange Wang, Chunyu Cao, Yanfeng Zhang, Chaoyi Chen, Hao Yuan, Yu Gu, Ge Yu</i>	
Rapidash: Efficient Detection of Constraint Violations.....	2009
<i>Zifan Liu, Shaleen Deep, Anna Fariha, Fotis Psallidas, Ashish Tiwari, Avrilia Floratou</i>	
Differentially Private Data Generation with Missing Data	2022
<i>Shubhankar Mohapatra, Jianqiao Zong, Florian Kerschbaum, Xi He</i>	
Everything You Always Wanted to Know About Storage Compressibility of Pre-Trained ML Models but Were Afraid to Ask.....	2036
<i>Zhaoyuan Su, Ammar Ahmed, Zirui Wang, Ali Anwar, Yue Cheng</i>	
Fight Fire with Fire: Towards Robust Graph Neural Networks on Dynamic Graphs via Actively Defense	2050
<i>Haoyang Li, Shimin Di, Calvin Hong Yi Li, Lei Chen, Xiaofang Zhou</i>	
SeLeP: Learning Based Semantic Prefetching for Exploratory Database Workloads	2064
<i>Farzaneh Zarak, Farhana Choudhury, Renata Borovica-Gajic</i>	
Contributions Estimation in Federated Learning: A Comprehensive Experimental Evaluation	2077
<i>Yiwei Chen, Kaiyu Li, Guoliang Li, Yong Wang</i>	
Visualization-aware Time Series Min-Max Caching with Error Bound Guarantees	2091
<i>Stavros Maroulis, Vassilis Stamatopoulos, George Papastefanatos, Manolis Terrovitis</i>	
CHORUS: Foundation Models for Unified Data Discovery and Exploration	2104
<i>Moe Kayali, Anton Lykov, Ilias Fountalis, Nikolaos Vasiloglou, Dan Olteanu, Dan Suciu</i>	
Extending Graph Rules with Oracles.....	2115
<i>Viktor Leis, Christian Dietrich</i>	

TABLE OF CONTENTS

Research Papers

CIVET: Exploring Compact Index for Variable-Length Subsequence Matching on Time Series	2123
<i>Haoran Xiong, Hang Zhang, Zeyu Wang, Zhenying He, Peng Wang, X. Sean Wang</i>	
Spatialyze: A Geospatial Video Analytics System with Spatial-Aware Optimizations	2136
<i>Chanwut Kittivorawong, Yongming Ge, Yousef Helal, Alvin Cheung</i>	
Optimal Matrix Sketching over Sliding Windows.....	2149
<i>Hanyan Yin, Dongxie Wen, Jiajun Li, Zhewei Wei, Xiao Zhang, Zengfeng Huang, Feifei Li</i>	
Window Function Expression: Let the Self-join Enter	2162
<i>Radim Baca</i>	
SplitDF: Splitting Dataframes for Memory-Efficient Data Analysis.....	2175
<i>Aarati Kakaraparthi, Jignesh M. Patel</i>	
Sampling Methods for Inner Product Sketching.....	2185
<i>Majid Daliri, Juliana Freire, Christopher Musco, Aécio Santos, Haoxiang Zhang</i>	
DIDS: Double Indices and Double Summarizations for Fast Similarity Search.....	2198
<i>Han Hu, Jiye Qiu, Hongzhi Wang, Bin Liang, Songling Zou</i>	
Improving Graph Compression for Efficient Resource-Constrained Graph Analytics	2212
<i>Qian Xu, Juan Yang, Feng Zhang, Zheng Chen, Jiawei Guan, Kang Chen, Ju Fan, Youren Shen, Ke Yang, Yu Zhang, Xiaoyong Du</i>	
Efficient Unsupervised Community Search with Pre-trained Graph Transformer	2227
<i>Jianwei Wang, Kai Wang, Xuemin Lin, Wenjie Zhang, Ying Zhang</i>	
DET-LSH: A Locality-Sensitive Hashing Scheme with Dynamic Encoding Tree for Approximate Nearest Neighbor Search.....	2241
<i>Jiuqi Wei, Botao Peng, Xiaodong Lee, Themis Palpanas</i>	
BIRD: Efficient Approximation of Bidirectional Hidden Personalized PageRank	2255
<i>Haoyu Liu, Siqiang Luo</i>	
GENTI: GPU-powered Walk-based Subgraph Extraction for Scalable Representation Learning on Dynamic Graphs.....	2269
<i>Zihao Yu, Ningyi Liao, Siqiang Luo</i>	
ArcheType: A Novel Framework for Open-Source Column Type Annotation using Large Language Models	2279
<i>Benjamin Feuer, Yurong Liu, Chinmay Hegde, Juliana Freire</i>	

Trajectory Similarity Measurement: An Efficiency Perspective	2293
<i>Yanchuan Chang, Egemen Tanin, Gao Cong, Christian S. Jensen, Jianzhong Qi</i>	
BYO: A Unified Framework for Benchmarking Large-Scale Graph Containers	2307
<i>Brian Wheatman, Xiaojun Dong, Zheqi Shen, Laxman Dhulipala, Jakub Łęcki, Prashant Pandey, Helen Xu</i>	
Secure and Verifiable Data Collaboration with Low-Cost Zero-Knowledge Proofs.....	2321
<i>Yizheng Zhu, Yuncheng Wu, Zhaojing Luo, Beng Chin Ooi, Xiaokui Xiao</i>	
Rashnu: Data-Dependent Order-Fairness	2335
<i>Heena Nagda, Shubhendra Pal Singh, Mohammad Javad Amiri, Boon Thau Loo</i>	
Sparcle: Boosting the Accuracy of Data Cleaning Systems through Spatial Awareness	2349
<i>Yuchuan Huang, Mohamed F. Mokbel</i>	
TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods.....	2363
<i>Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Linn Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, Zhenli Sheng, Bin Yang</i>	
FSM: A Fine-grained Splitting and Merging Framework for Dual-balanced Graph Partition.....	2378
<i>Chengjun Liu, Zhuo Peng, Weiguo Zheng, Lei Zou</i>	
Efficient and Reliable Estimation of Knowledge Graph Accuracy	2392
<i>Stefano Marchesin, Gianmaria Silvello</i>	

TABLE OF CONTENTS

Research Papers

Breaking It Down: An In-depth Study of Index Advisors	2405
<i>Wei Zhou, Chen Lin, Xuanhe Zhou, Guoliang Li</i>	
Accelerating Maximal Clique Enumeration via Graph Reduction	2419
<i>Wen Deng, Weiguo Zheng, Hong Cheng</i>	
POLIGRAS: Policy-based Graph Summarization	2432
<i>Jiyang Bai, Peixiang Zhao</i>	
SWAT: A System-Wide Approach to Tunable Leakage Mitigation in Encrypted Data Stores	2445
<i>Leqian Zheng, Lei Xu, Cong Wang, Sheng Wang, Yuke Hu, Zhan Qin, Feifei Li, Kui Ren</i>	
TIGER: Training Inductive Graph Neural Network for Large-scale Knowledge Graph Reasoning	2459
<i>Kai Wang, Yuwei Xu, Siqiang Luo</i>	
Incremental Sliding Window Connectivity over Streaming Graphs	2473
<i>Chao Zhang, Angela Bonifati, M. Tamer Özsu</i>	
CohortNet: Empowering Cohort Discovery for Interpretable Healthcare Analytics	2487
<i>Qingpeng Cai, Kaiping Zheng, H. V. Jagadish, Beng Chin Ooi, James Yip</i>	
Efficient Influence Minimization via Node Blocking	2501
<i>Jiansen Song, Yanping Wu, Xiaoyang Wang, Ying Zhang, Lu Qin, Wenjie Zhang, Xuemin Lin</i>	
D-Bot: Database Diagnosis System using Large Language Models	2514
<i>Xuanhe Zhou, Guoliang Li, Zhaoyan Sun, Zhiyuan Liu, Weize Chen, Jianming Wu, Jiesi Liu, Ruohang Feng, Guoyang Zeng</i>	
Blitzcrank: Fast Semantic Compression for In-memory Online Transaction Processing	2528
<i>Yiming Qiao, Yihan Gao, Huanchen Zhang</i>	
Spectrum: Speedy and Strictly-Deterministic Smart Contract Transactions for Blockchain Ledgers	2541
<i>Zhihao Chen, Tianji Yang, Yixiao Zheng, Zhao Zhang, Cheqing Jin, Aoying Zhou</i>	
Fast Commitment for Geo-Distributed Transactions via Decentralized Co-coordinators	2555
<i>Zihao Zhang, Huiqi Hu, Xuan Zhou, Yaofeng Tu, Weining Qian, Aoying Zhou</i>	
CXL and the Return of Scale-Up Database Engines	2568
<i>Alberto Lerner, Gustavo Alonso</i>	
Inductive Attributed Community Search: to Learn Communities across Graphs	2576
<i>Shuheng Fang, Kangfei Zhao, Yu Rong, Zhixun Li, Jeffrey Xu Yu</i>	
I/O Efficient Label-Constrained Reachability Queries in Large Graphs	2590
<i>Long Yuan, Xia Li, Zi Chen, Xuemin Lin, Xiang Zhao, Wenjie Zhang</i>	

DEX: Scalable Range Indexing on Disaggregated Memory	2603
<i>Baotong Lu, Kaisong Huang, Chieh-Jan Mike Liang, Tianzheng Wang, Eric Lo</i>	
Automatic Data Repair: Are We Ready to Deploy?	2617
<i>Wei Ni, Xiaoye Miao, Xiangyu Zhao, Yangyang Wu, Shuwei Liang, Jianwei Yin</i>	
Biathlon: Harnessing Model Resilience for Accelerating ML Inference Pipelines	2631
<i>Chaokun Chang, Eric Lo, Chunxiao Ye,</i>	
Distributed Shortest Distance Labeling on Large-Scale Graphs	2641
<i>Yuanyuan Zeng, Chenhao Ma, Yixiang Fang</i>	
Efficient Parallel D-core Decomposition at Scale	2654
<i>Wensheng Luo, Yixiang Fang, Chunxu Lin, Yingli Zhou</i>	
Efficient Discovery of Significant Patterns with Few-Shot Resampling	2668
<i>Leonardo Pellegrina, Fabio Vandin</i>	

TABLE OF CONTENTS

Research Papers

Robust Best Point Selection under Unreliable User Feedback.....	2681
<i>Qixu Chen, Raymond Chi-Wing Wong</i>	
Towards Optimal Transaction Scheduling.....	2694
<i>Audrey Cheng, Aaron Kabcenell, Jason Chan, Xiao Shi, Peter Bailis, Natacha Crooks, Ion Stoica</i>	
QCore: Data-Efficient, On-Device Continual Calibration for Quantized Models.....	2708
<i>David Campos, Bin Yang, Tung Kieu, Miao Zhang, Chenjuan Guo, Christian S. Jensen</i>	
Efficient Algorithms for Pseudoarboricity Computation in Large Static and Dynamic Graphs.....	2722
<i>Yalong Zhang, Ronghua Li, Qi Zhang, Hongchao Qin, Lu Qin, Guoren Wang</i>	
RoarGraph: A Projected Bipartite Graph for Efficient Cross-Modal Approximate Nearest Neighbor Search.....	2735
<i>Meng Chen, Kai Zhang, Zhenying He, Yinan Jing, X. Sean Wang</i>	
Combining Small Language Models and Large Language Models for Zero-Shot NL2SQL.....	2750
<i>Ju Fan, Zihui Gu, Songyue Zhang, Yuxin Zhang, Zui Chen, Lei Cao, Guoliang Li, Samuel Madden, Xiaoyong Du, Nan Tang</i>	
D3-GNN: Dynamic Distributed Dataflow for Streaming Graph Neural Networks.....	2764
<i>Rustam Guliyev, Aparajita Haldar, Hakan Ferhatosmanoglu</i>	
Distance-based Outlier Query Optimization in Apache IoTDB.....	2778
<i>Yunxiang Su, Shaoxu Song, Xiangdong Huang, Chen Wang, Jianmin Wang</i>	
TC-Match: Fast Time-constrained Continuous Subgraph Matching.....	2791
<i>Jianye Yang, Sheng Fang, Zhaoquan Gu, Ziyi Ma, Xuemin Lin, Zhihong Tian</i>	
Automating the Enterprise with Foundation Models.....	2805
<i>Michael Wornow, Avanika Narayan, Krista Opsahl-Ong, Quinn McIntyre, Nigam Shah, Christopher Re</i>	
Efficient Index for Temporal Core Queries over Bipartite Graphs.....	2813
<i>Anxin Tian, Alexander Zhou, Yue Wang, Xun Jian, Lei Chen</i>	
Uldp-FL: Federated Learning with Across Silo User-Level Differential Privacy.....	2826
<i>Fumiyuki Kato, Li Xiong, Shun Takagi, Yang Cao, Masatoshi Yoshikawa</i>	
Evolution Forest Index: Towards Optimal Temporal k -Core Component Search via Time-Topology Isomorphic Computation.....	2840
<i>Junyong Yang, Ming Zhong, Yuanyuan Zhu, Tieyun Qian, Mengchi Liu, Jeffrey Xu Yu</i>	

Eliminating Data Processing Bottlenecks in GNN Training over Large Graphs via Two-level Feature Compression	2854
<i>Yuxin Ma, Ping Gong, Tianming Wu, Jiawei Yi, Chengru Yang, Cheng Li, Qirong Peng, Guiming Xie, Yongcheng Bao, Haifeng Liu, Yinlong Xu</i>	
Towards Systematic Index Dynamization	2867
<i>Douglas B. Rumbaugh, Dong Xie, Zhuoyue Zhao</i>	
Ensemble Clustering based on Meta-Learning and Hyperparameter Optimization	2880
<i>Dennis Treder-Tschechlov, Manuel Fritz, Holger Schwarz, Bernhard Mitschang</i>	
Efficient Stochastic Routing in Path-Centric Uncertain Road Networks.....	2893
<i>Chenjuan Guo, Ronghui Xu, Bin Yang, Yuan Ye, Tung Kieu, Yan Zhao, Christian S. Jensen</i>	
Transforming Property Graphs	2906
<i>Angela Bonifati, Filip Murlak, Yann Ramusat</i>	
Are Large Language Models a Good Replacement of Taxonomies?	2919
<i>Yushi Sun, Xin Hao, Kai Sun, Yifan Ethan Xu, Xiao Yang, Xin Luna Dong, Nan Tang, Lei Chen</i>	
Efficient Algorithms for Density Decomposition on Large Static and Dynamic Graphs	2933
<i>Yalong Zhang, Ronghua Li, Qi Zhang, Hongchao Qin, Guoren Wang</i>	
Efficient Maximal Motif-Clique Enumeration over Large Heterogeneous Information Networks	2946
<i>Yingli Zhou, Yixiang Fang, Chenhao Ma, Tianci Hou, Xin Huang</i>	
OUTRE: An OUT-of-core De-REdundancy GNN Training Framework for Massive Graphs within A Single Machine.....	2960
<i>Zeang Sheng, Wentao Zhang, Yangyu Tao, Bin Cui</i>	
On Reducing Space Amplification with Multi-Column Compaction in Apache IoTDB.....	2974
<i>Chenguang Fang, Zijie Chen, Shaoxu Song, Xiangdong Huang, Chen Wang, Jianmin Wang</i>	
AutoTSAD: Unsupervised Holistic Anomaly Detection for Time Series Data	2987
<i>Sebastian Schmidl, Felix Naumann, Thorsten Papenbrock</i>	
Enabling Window-Based Monotonic Graph Analytics with Reusable Transitional Results for Pattern-Consistent Queries	3003
<i>Zheng Chen, Feng Zhang, Yang Chen, Xiaokun Fang, Guanyu Feng, Xiaowei Zhu, Wenguang Chen, Xiaoyong Du</i>	
When Amnesia Strikes: Understanding and Reproducing Data Loss Bugs with Fault Injection	3017
<i>Maria Ramos, João Azevedo, Kyle Kingsbury, José Pereira, Tânia Esteves, Ricardo Macedo, João Paulo</i>	
PriPL-Tree: Accurate Range Query for Arbitrary Distribution under Local Differential Privacy	3031
<i>Leixia Wang, Qingqing Ye, Haibo Hu, Xiaofeng Meng</i>	
Win-Win: On Simultaneous Clustering and Imputing over Incomplete Data	3045
<i>Yu Sun, Jingyu Zhu, Xiao Xu, Xian Xu, Yuyao Sun, Shaoxu Song, Xiang Li, Xiaojie Yuan</i>	
HRNet: Differentially Private Hierarchical and Multi-Resolution Network for Human Mobility Data Synthesization.....	3058
<i>Shun Takagi, Li Xiong, Fumiyuki Kato, Yang Cao, Masatoshi Yoshikawa</i>	

Efficiently Mitigating the Impact of Data Drift on Machine Learning Pipelines	3072
<i>Sijie Dong, Qitong Wang, Soror Sahri, Themis Palpanas, Divesh Srivastava</i>	
PCSP: Efficiently Answering Label-Constrained Shortest Path Queries in Road Networks.....	3082
<i>Libin Wang, Raymond Chi-Wing Wong</i>	
Efficient Enumeration of Recursive Plans in Transformation-based Query Optimizers.....	3095
<i>Amela Fejza, Pierre Genevès, Nabil Layaida</i>	
Enriching Relations with Additional Attributes for ER	3109
<i>Mengyi Yan, Wenfei Fan, Yaoshu Wang, Min Xie</i>	
Enhancing Accuracy for Super Spreader Identification in High-Speed Data Streams.....	3124
<i>Haibo Wang</i>	
Privately Answering Queries on Skewed Data via Per-Record Differential Privacy	3138
<i>Jeremy Seeman, William Sexton, David Pujol, Ashwin Machanavajhala</i>	
Agile-Ant: Self-managing Distributed Cache Management for Cost Optimization of Big Data Applications	3151
<i>Hani Al-Sayeh, Muhammad Attahir Jibril, Kai-Uwe Sattler</i>	
Complex Event Recognition with Symbolic Register Transducers	3165
<i>Elias Alevizos, Alexander Artikis, Georgios Paliouras</i>	
HAChart: Human and AI Paired Visualization System	3178
<i>Yupeng Xie, Yuyu Luo, Guoliang Li, Nan Tang</i>	
A Sampling-based Framework for Hypothesis Testing on Large Attributed Graphs.....	3192
<i>Yun Wang, Chrysanthi Kosyfaki, Sihem Amer-Yahia, Reynold Cheng</i>	
LLM-PBE: Assessing Data Privacy in Large Language Models	3201
<i>Qinbin Li, Junyuan Hong, Chulin Xie, Jeffrey Tan, Rachel Xin, Junyi Hou, Xavier Yin, Zhun Wang, Dan Hendrycks, Zhangyang Wang, Bo Li, Bingsheng He, Dawn Song</i>	
Robust Join Processing with Diamond Hardened Joins.....	3215
<i>Altan Birler, Alfons Kemper, Thomas Neumann</i>	
DARKER: Efficient Transformer with Data-driven Attention Mechanism for Time Series	3229
<i>Rundong Zuo, Guozhong Li, Rui Cao, Byron Choi, Jianliang Xu, Sourav S Bhowmick</i>	
Efficient Maximal Frequent Group Enumeration in Temporal Bipartite Graphs.....	3243
<i>Yanping Wu, Renjie Sun, Xiaoyang Wang, Dong Wen, Ying Zhang, Lu Qin, Xuemin Lin</i>	
Optimizing Video Queries with Declarative Clues	3256
<i>Daren Chao, Yueting Chen, Nick Koudas, Xiaohui Yu</i>	
Fainder: A Fast and Accurate Index for Distribution-Aware Dataset Search	3269
<i>Lennart Behme, Sainyam Galhotra, Kaustubh Beedkar, Volker Markl</i>	
nsDB: Architecting the Next Generation Database by Integrating Neural and Symbolic Systems.....	3283
<i>Ye Yuan, Bo Tang, Tianfei Zhou, Zhiwei Zhang, Jianbin Qin</i>	
Two Birds With One Stone: Designing a Hybrid Cloud Storage Engine for HTAP	3290
<i>Tobias Schmidt, Dominik Durner, Viktor Leis, Thomas Neumann</i>	

DDS: DPU-optimized Disaggregated Storage	3304
<i>Qizhen Zhang, Philip A. Bernstein, Badrish Chandramouli, Jiasheng Hu, Yiming Zheng</i>	
The Dawn of Natural Language to SQL: Are We Fully Ready?	3318
<i>Boyan Li, Yuyu Luo, Chengliang Chai, Guoliang Li, Nan Tang</i>	
Counterfactual Explanation of the Shapley Value in Data Coalitions	3332
<i>Michelle Si, Jian Pei</i>	
Searching Data Lakes for Nested and Joined Data	3346
<i>Yi Zhang, Peter Baile Chen, Zackary G. Ives</i>	
Efficient Betweenness Centrality Computation over Large Heterogeneous Information Networks	3360
<i>Xinrui Wang, Wang Yiran, Xuemin Lin, Jeffrey Xu Yu, Hong Gao, Xiuzhen Cheng, Dongxiao Yu</i>	
The Holon Approach for Simultaneously Tuning Multiple Components in a Self-Driving Database Management System with Machine Learning via Synthesized Proto-Actions	3373
<i>William Zhang, Wan Shen Lim, Matthew Butrovich, Andrew Pavlo</i>	
DynaHB: A Communication-Avoiding Asynchronous Distributed Framework with Hybrid Batches for Dynamic GNN Training	3388
<i>Zhen Song, Yu Gu, Qing Sun, Tianyi Li, Yanfeng Zhang, Yushuai Li, Christian S. Jensen, Ge Yu</i>	
Efficient kNN Search in Public Transportation Networks	3402
<i>Qingshuai Feng, Junhua Zhang, Wenjie Zhang, Lu Qin, Ying Zhang, Xuemin Lin</i>	
LITS: An Optimized Learned Index for Strings	3415
<i>Yifan Yang, Shimin Chen</i>	
OLAP on Modern Chiplet-Based Processors	3428
<i>Alessandro Fogli, Bo Zhao, Peter Pietzuch, Maximilian Bandle, Jana Giceva</i>	
Bf-Tree: A Modern Read-Write-Optimized Concurrent Larger-Than-Memory Range Index	3442
<i>Xiangpeng Hao, Badrish Chandramouli</i>	
Partition, Don't Sort! Compression Boosters for Cloud Data Ingestion Pipelines	3456
<i>Patrick Hansert, Sebastian Michel</i>	
Chameleon: Foundation Models for Fairness-aware Multi-modal Data Augmentation to Enhance Coverage of Minorities	3470
<i>Mahdi Erfanian, H. V. Jagadish, Abolfazl Asudeh</i>	
DAFDiscover: Robust Mining Algorithm for Dynamic Approximate Functional Dependencies on Dirty Data	3484
<i>Xiaoou Ding, Yixing Lu, Hongzhi Wang, Chen Wang, Yida Liu, Jianmin Wang</i>	
Hardware-Efficient Data Imputation through DBMS Extensibility	3497
<i>Hubert Mohr-Daurat, Georgios Theodorakis, Holger Pirk</i>	
Generating Succinct Descriptions of Database Schemata for Cost-Efficient Prompting of Large Language Models	3511
<i>Immanuel Trummer</i>	
Saving Money for Analytical Workloads in the Cloud	3524
<i>Tapan Srivastava, Raul Castro Fernandez</i>	

ReCG: Bottom-Up JSON Schema Discovery Using a Repetitive Cluster-and-Generalize Framework	3538
<i>Joohyung Yun, Byungchul Tak, Wook-Shin Han</i>	
Optimizing Collections of Bloom Filters within a Space Budget.....	3551
<i>Gabriel Mersy, Zhuo Wang, Stavros Sintos, Sanjay Krishnan</i>	
A Spark Optimizer for Adaptive, Fine-Grained Parameter Tuning	3565
<i>Chenghao Lyu, Qi Fan, Philippe Guyard, Yanlei Diao</i>	
Texera: A System for Collaborative and Interactive Data Analytics Using Workflows	3580
<i>Zuozhi Wang, Yicong Huang, Shengquan Ni, Avinash Kumar, Sadeem Alsudais, Xiaozhen Liu, Xinyuan Lin, Yunyan Ding, Chen Li</i>	
Efficient Validation of SHACL Shapes with Reasoning.....	3589
<i>Jin Ke, Zenon Zacouris, Maribel Acosta</i>	
QED: A Powerful Query Equivalence Decider for SQL	3602
<i>Shuxian Wang, Sicheng Pan, Alvin Cheung</i>	
Index Advisors on Quantum Platforms.....	3615
<i>Manish Kesarwani, Jayant R. Haritsa</i>	
Blueprinting the Cloud: Unifying and Automatically Optimizing Cloud Data Infrastructures with BRAD	3629
<i>Geoffrey X. Yu, Ziniu Wu, Ferdi Kossmann, Tianyu Li, Markos Markakis, Amadou Ngom, Samuel Madden, Tim Kraska</i>	
Aleph Filter: To Infinity in Constant Time.....	3644
<i>Niv Dayan, Ioana-Oriana Bercea, Rasmus Pagh</i>	
RUSH: Real-time Burst Subgraph Discovery in Dynamic Graphs.....	3657
<i>Yuhang Chen, Jiaxin Jiang, Shixuan Sun, Bingsheng He, Min Chen</i>	
A Benchmark Study of Deep-RL Methods for Maximum Coverage Problems over Graphs	3666
<i>Zhicheng Liang, Yu Yang, Xiangyu Ke, Xiaokui Xiao, Yunjun Gao</i>	
Hit the Gym: Accelerating Query Execution to Efficiently Bootstrap Behavior Models for Self-Driving Database Management Systems.....	3680
<i>Wan Shen Lim, Lin Ma, William Zhang, Matthew Butrovich, Samuel Arch, Andrew Pavlo</i>	
Why TPC Is Not Enough: An Analysis of the Amazon Redshift Fleet.....	3694
<i>Alexander van Renen, Dominik Horn, Pascal Pfeil, Kapil Vaidya, Wenjian Dong, Murali Narayanaswamy, Zhengchun Liu, Gaurav Saxena, Andreas Kipf, Tim Kraska</i>	
Efficient k -Clique Count Estimation with Accuracy Guarantee.....	3707
<i>Lijun Chang, Rashmika Gamage, Jeffrey Xu Yu</i>	

TABLE OF CONTENTS

Industrial Papers

Cloud Actor-Oriented Database Transactions in Orleans.....	3720
<i>Tamer Eldeeb, Sebastian Burckhardt, Reuben Bond, Asaf Cidon, Junfeng Yang, Philip A. Bernstein</i>	
ClickHouse - Lightning Fast Analytics for Everyone	3731
<i>Robert Schulze, Tom Schreiber, Ilya Yatsishin, Ryadh Dahimene, Alexey Milovidov</i>	
PALF: Replicated Write-ahead Logging for Distributed Databases	3745
<i>Fusheng Han, Hao Liu, Bin Chen, Debin Jia, Jianfeng Zhou, Xuwang Teng, Chuanhui Yang, Huafeng Xi, Wei Tian, Shuning Tao, Sen Wang, Quanqing Xu, Zhenkun Yang</i>	
Towards Resource Efficiency: Practical Insights into Large-Scale Spark Workloads at ByteDance	3759
<i>Yixin Wu, Xiuqi Huang, Wei Zhongjia, Hang Cheng, Chaohui Xin, Zuzhi Chen, Binbin Chen, Yufei Wu, Hao Wang, Tieying Zhang, Rui Shi, Xiaofeng Gao, Yuming Liang, Pengwei Zhao, Guihai Chen</i>	
SingleStore-V: An Integrated Vector Database System in SingleStore.....	3772
<i>Cheng Chen, Chenzhe Jin, Yunan Zhang, Sasha Podolsky, Chun Wu, Szu-Po Wang, Eric Hanson, Zhou Sun, Robert Walzer, Jianguo Wang</i>	
GaussDB: A Cloud-Native Multi-Primary Database with Compute-Memory-Storage Disaggregation	3786
<i>Guoliang Li, Wengang Tian, Jinyu Zhang, Ronen Grosman, Zongchao Liu, Sihao Li</i>	
LavaStore: ByteDance's Purpose-built, High-performance, Cost-effective Local Storage Engine for Cloud Services	3799
<i>Hao Wang, Jiaxin Ou, Ming Zhao, Sheng Qiu, Yizheng Jiao, Yi Wang, Qizhong Mao, Zhengyu Yang, Yang Liu, Jianyang Hu, Jingwei Zhang, Jinrui Liu, Jiaqiang Chen, Yong Sheng, Cao Lixun, Heng Zhang, Hongde Li, Ming Li, Yue Ma, Lei Zhang, Jian Liu, Guanghui Zhang, Fei Liu, Jianjun Chen</i>	
Membrane - Safe and Performant Data Access Controls in Apache Spark in the Presence of Imperative Code	3813
<i>Andrei Paduroiu, Sungheun Wi, Yan Yan, Roni Burd, Ruhollah Farchtchi, Giovanni Matteo Fumarola</i>	
An Examination of CXL Memory Use Cases for In-Memory Database Management Systems using SAP HANA	3827
<i>Minseon Ahn, Thomas Willhalm, Norman May, Donghun Lee, Suprasad Mutalik Desai, Daniel Booss, Jungmin Kim, Navneet Singh, Daniel Ritter, Oliver Rebholz</i>	
KGFabric: A Scalable Knowledge Graph Warehouse for Enterprise Data Interconnection	3841
<i>Peng Yi, Lei Liang, Zhang Da, Chen Yong, Jinye Zhu, Xiangyu Liu, Kun Tang, Jialin Chen, Hao Lin, Leijie Qiu, Jun Zhou</i>	

Db2une: Tuning Under Pressure via Deep Learning	3855
<i>Alexander Bianchi, Andrew Chai, Vincent Corvinelli, Parke Godfrey, Jarek Szlichta, Calisto Zuzarte</i>	
TDSQL: Tencent Distributed Database System	3869
<i>Yuxing Chen, Anqun Pan, Hailin Lei, Anda Ye, Shuo Han, Yan Tang, Wei Lu, Yunpeng Chai, Feng Zhang, Xiaoyong Du</i>	
A Flexible Forecasting Stack	3883
<i>Tim Januschowski, Yuyang Wang, Jan Gasthaus, Syama Rangapuram, Caner Turkmen, Jasper Zschiegner, Lorenzo Stella, Michael Bohlke-Schneider, Danielle Maddix, Konstantinos Benidis, Alexander Alexandrov, Christos Faloutsos, Sebastian Schelter</i>	
Galaxybase: A High Performance Native Distributed Graph Database for HTAP	3893
<i>Bing Tong, Yan Zhou, Chen Zhang, Jianheng Tang, Jing Tang, Leihong Yang, Qiye Li, Manwu Lin, Zhongxin Bao, Jia Li, Lei Chen</i>	
SecuDB: An In-enclave Privacy-preserving and Tamper-resistant Relational Database	3906
<i>Xinying Yang, Cong Yue, Wenhui Zhang, Yang Liu, Beng Chin Ooi, Jianjun Chen</i>	
AutoTQA: Towards Autonomous Tabular Question Answering through Multi-Agent Large Language Models	3920
<i>Jun-Peng Zhu, Peng Cai, Kai Xu, Li Li, Yishen Sun, Shuai Zhou, Haihuang Su, Liu Tang, Qi Liu</i>	
ResLake: Towards Minimum Job Latency and Balanced Resource Utilization in Geo-distributed Job Scheduling.....	3934
<i>XinChun Zhang, Aqsa Kashaf, Yihan Zou, Wei Zhang, Weibo Liao, Song Haoxiang, Jintao Ye, Yakun Li, Rui Shi, Yong Tian, Feng Wei, Binbin Chen, Zuzhi Chen, Tieying Zhang, Yongping Tang</i>	
Adaptive and Robust Query Execution for Lakehouses At Scale.....	3947
<i>Maryann Xue, Yingyi Bu, Abhishek Somani, Wenchen Fan, Ziqi Liu, Steven Chen, Herman Van Hovell, Bart Samwel, Mostafa Mokhtar, Rk Korlapati, Andy Lam, Yunxiao Ma, Vuk Ercegovic, Jiexing Li, Alexander Behm, Yuanjian Li, Xiao Li, Sriram Krishnamurthy, Amit Shukla, Michalis Petropoulos, Sameer Paranjpye, Reynold Xin, Matei Zaharia</i>	
Transparent Migration from Datastore to Firestore	3960
<i>Ed Davisson, Tilo Dickopp, David Gay, Eric Karasuda, Ram Kesavan, Vadim Yushprakh</i>	
Complex-Path: Effective and Efficient Node Ranking with Paths in Billion-Scale Heterogeneous Graphs	3973
<i>Jinquan Hang, Zhiqing Hong, Xinyue Feng, Guang Wang, Dongjiang Cao, Jiayang Qiao, Haotian Wang, Desheng Zhang</i>	
SecretFlow-SCQL: A Secure Collaborative Query pLatform	3987
<i>Wenjing Fang, Shunde Cao, Guojin Hua, Junming Ma, Yongqiang Yu, Qunshan Huang, Jun Feng, Jin Tan, Xiaopeng Zan, Pu Duan, Yang Yang, Li Wang, Ke Zhang, Lei Wang</i>	
Towards Millions of Database Transmission Services in the Cloud	4001
<i>Hua Fan, Dachao Fu, Xu Wang, Jiachi Zhang, Chaoji Zuo, Zhengyi Wu, Miao Zhang, Kang Yuan, Xizi Ni, Huo Guocheng, Wenchao Zhou, Feifei Li, Jingren Zhou</i>	
Large-Scale Metric Computation in Online Controlled Experiment Platform	4014
<i>Tao Xiong, Yong Wang</i>	
X-Stor: A Cloud-native NoSQL Database Service with Multi-model Support.....	4025
<i>Hongyu Lei, Chunhua Li, Ke Zhou, Jianping Zhu, Kezhou Yan, Fen Xiao, Ming Xie, Jiang Wang, Shiyu Di</i>	

Resource Management in Aurora Serverless	4038
<i>Bradley Barnhart, Marc Brooker, Daniil Chinenkov, Tony Hooper, Jihoun Im, Prakash Chandra Jha, Tim Kraska, Ashok Kurakula, Alexey Kuznetsov, Grant Mcalister, Arjun Muthukrishnan, Aravinthan Narayanan, Douglas Terry, Bhuvan Uргаonkar, Jiaming Yan</i>	
SQL has problems. We can fix them: Pipe syntax in SQL	4051
<i>Jeff Shute, Shannon Bales, Matthew Brown, Jean-Daniel Browne, Brandon Dolphin, Romit Kuditarkar, Andrey Litvinov, Jingchi Ma, John D Morcos, Michael Shen, David Wilhite, Xi Wu, Lulan Yu</i>	
Apache TsFile: An IoT-native Time Series File Format	4064
<i>Xin Zhao, Jialin Qiao, Xiangdong Huang, Chen Wang, Shaoxu Song, Jianmin Wang</i>	
Presto's History-based Query Optimizer	4077
<i>Pranjal Shankhdhar, Feilong Liu, Jay Narale, James Sun, Rebecca Schlussel, Lyublena Antova</i>	
OptScaler: A Collaborative Framework for Robust Autoscaling in the Cloud	4090
<i>Ding Zou, Wei Lu, Zhibo Zhu, Xingyu Lu, Jun Zhou, Xiaojin Wang, Kangyu Liu, Kefan Wang, Renen Sun, Haiqing Wang</i>	
Dealing with Acronyms, Abbreviations, and Typos in Real-World Entity Matching.....	4104
<i>Joshua Wu, Dixin Tang, Nithin V Chalapathi, Tristan Chambers, Julie Ciccolini, Cheryl Phillips, Lisa Pickoff-White, Aditya Parameswaran</i>	
Lindorm-UWC: An Ultra-Wide-Column Database for Internet of Vehicles.....	4117
<i>Qianyu Ouyang, Chunhui Shen, Wenlong Yang, Peng Yu, Qiang Xiao, Jianhui Lei, Yadong Chen, Qilu Zhong, Xiang Wang, Yong Lin, Qingyi Meng, Zhicheng Ji, Wei Meng, Cen Zheng, Sheng Wang, Dan Pei, Wei Zhang, Feifei Li, Jingren Zhou</i>	
DLRover-RM: Resource Optimization for Deep Recommendation Models Training in the Cloud.....	4130
<i>Qinlong Wang, Tingfeng Lan, Yinghao Tang, Bo Sang, Ziling Huang, Yiheng Du, Haitao Zhang, Jian Sha, Hui Lu, Yuanchun Zhou, Ke Zhang, Mingjie Tang</i>	
Differentially Private Stream Processing at Scale	4145
<i>Bing Zhang, Vadym Doroshenko, Peter Kairouz, Thomas Steinke, Abhradeep Thakurta, Ziyin Ma, Eidan Cohen, Himani Apte, Jodi Spacek</i>	
Petabyte-Scale Row-Level Operations in Data Lakehouses.....	4159
<i>Anton Okolnychyi, Chao Sun, Kazuyuki Tanimura, Russell Spitzer, Ryan Blue, Szehon Ho, Yufei Gu, Vishwanath Lakkundi, Db Tsai</i>	
SPADE: Synthesizing Data Quality Assertions for Large Language Model Pipelines	4173
<i>Shreya Shankar, Haotian Li, Parth Asawa, Madelon Hulsebos, Yiming Lin, J.D. Zamfirscu-Pereira, Harrison Chase, William Fu-Hinthorn, Aditya Parameswaran, Eugene Wu</i>	
Simple (yet Efficient) Function Authoring for Vectorized Engines	4187
<i>Laith Sakka, Pedro Pedreira, Orri Erling, Masha Basmanova, Kevin Wilfong, Wei He, Xiaoxuan Meng, Krishna Pai, Bikramjeet Vig</i>	
Grouping, Subsumption, and Disjunctive Join Optimizations in Oracle	4200
<i>Rafi Ahmed, Krishna Kantikiran Pasupuleti, Sriram Tirupattur, Lei Sheng, Hong Su, Mohamed Ziauddin</i>	

Tutorials

LLM for Data Management	4213
<i>Guoliang Li, Xuanhe Zhou, Xinyang Zhao</i>	
Native Distributed Databases: Problems, Challenges and Opportunities	4217
<i>Quanqing Xu, Chuanhui Yang, Aoying Zhou</i>	
A Reproducible Tutorial on Reproducibility in Database Systems Research	4221
<i>Tim Fischer, Denis Hirn, Gokhan Kul</i>	
Fairness in Preference Queries: Social Choice Theories Meet Data Management.....	4225
<i>Senjuti Basu Roy, Baruch Schieber, Nimrod Talmon</i>	
Time-Series Anomaly Detection: Overview and New Trends	4229
<i>Qinghua Liu, Paul Boniol, Themis Palpanas, John Paparrizos</i>	
Consensus in Data Management With Use Cases in Edge-Cloud and Blockchain Systems	4233
<i>Faisal Nawab, Mohammad Sadoghi</i>	
Efficient Training of Graph Neural Networks on Large Graphs.....	4237
<i>Yanyan Shen, Lei Chen, Jingzhi Fang, Xin Zhang, Shihong Gao, Hongbo Yin</i>	
Workload Placement on Heterogeneous CPU-GPU Systems.....	4241
<i>Marcos N. L. Carvalho, Alkis Simitsis, Anna Queralt, Oscar Romero</i>	
Spatial Query Optimization With Learning	4245
<i>Xin Zhang, Ahmed Eldawy</i>	
Composable Data Management: An Execution Overview.....	4249
<i>Pedro Pedreira, Deepak Majeti, Orri Erling</i>	

Demonstrations

Spade: A Real-Time Fraud Detection Framework	4253
<i>Jiaxin Jiang, Zhen Zhang, Bingqiao Luo, Bingsheng He, Min Chen, Wei Yang Wang, Jia Chen</i>	
A Data-driven Spatiotemporal Simulator for Reinforcement Learning Methods	4257
<i>Dingyuan Shi, Bingchen Song, Yuanyuan Zhang, Haolong Yang, Ke Xu</i>	
BFTGym: An Interactive Playground for BFT Protocols	4261
<i>Haoyun Qin, Chenyuan Wu, Mohammad Javad Amiri, Ryan Marcus, Boon Thau Loo</i>	
DTGraph: Declarative Transformations of Property Graphs	4265
<i>Angela Bonifati, Yann Ramusat, Filip Murlak, Amela Fejza, Rachid Echahed</i>	
MLOS in Action: Bridging the Gap Between Experimentation and Auto-Tuning in the Cloud	4269
<i>Brian Kroth, Sergiy Matushevych, Rana Alotaibi, Yiwen Zhu, Anja Gruenheid, Yuanyuan Tian</i>	
Snapcase - Regain Control over Your Predictions with Low-Latency Machine Unlearning.....	4273
<i>Sebastian Schelter, Stefan Grafberger, Maarten de Rijke</i>	

DiversiNews: Enriching News Consumption with Relevant yet Diverse News Articles Retrieval.....	4277
<i>Yiqun Sun, Qiang Huang, Yanhao Wang, Anthony K. H. Tung</i>	
SpannerLib: Embedding Declarative Information Extraction in an Imperative Workflow	4281
<i>Dean Light, Ahmad Aiashi, Mahmoud Diab, Daniel Nachmias, Stijn Vansummeren, Benny Kimelfeld</i>	
Demonstrating TabEE: Tabular Embedding Explanations	4285
<i>Roni Copul, Nave Frost, Tova Milo, Kathy Razmadze</i>	
Navigating Data Repositories: Utilizing Line Charts to Discover Relevant Datasets.....	4289
<i>Daomin Ji, Hui Luo, Zhifeng Bao, Shane Culpepper</i>	
Graph Association Analyses for Early Drug Discovery	4293
<i>Wenfei Fan, Daji Li, Peiyu Liang, Shuhao Liu, Yaoshu Wang, Yiming Wang, Min Xie, Runjie Zhang</i>	
Demonstration of MaskSearch: Efficiently Querying Image Masks for Machine Learning Workflows	4297
<i>Linsey Linxi Wei, Chung Yik Edward Yeung, Hongjian Yu, Jingchuan Zhou, Dong He, Magdalena Balazinska</i>	
Looking Deeply into the Magic Mirror: An Interactive Analysis of Database Index Selection Approaches	4301
<i>Stefan Halfpap, Jan Kossmann, Rainer Schlosser, Volker Markl</i>	
UTOPIA: Automatic Pivot Table Assistant.....	4305
<i>Whanhee Cho, Anna Fariha</i>	
TSGAssist: An Interactive Assistant Harnessing LLMs and RAG for Time Series Generation Recommendations and Benchmarking.....	4309
<i>Yihao Ang, Yifan Bao, Qiang Huang, Anthony K. H. Tung, Zhiyong Huang</i>	
DoppelGanger++ in Action: A Database Replay System with Fast Dependency Graph Generation	4313
<i>Wonseok Lee, Jaehyun Ha, Wook-Shin Han, Changgyoo Park, Myunggon Park, Juhyeng Han</i>	
LucidScript: Bottom-up Standardization for Data Preparation.....	4317
<i>Eugenie Y. Lai, Yuze Lou, Brit Youngmann, Michael Cafarella</i>	
OSSInsight: Scalable GitHub Analysis	4321
<i>Ahmad Ghazal, Zhiyuan Liang, Sunny Bains, Hanumath Maduri</i>	
IsoVista: Black-box Checking Database Isolation Guarantees.....	4325
<i>Long Gu, Si Liu, Tiancheng Xing, Hengfeng Wei, Yuxing Chen, David A Basin</i>	
ImputeVIS: An Interactive Evaluator to Benchmark Imputation Techniques for Time Series Data	4329
<i>Mourad Khayati, Quentin Nater, Jacques Pasquier</i>	
An Interactive Multi-modal Query Answering System with Retrieval-Augmented Large Language Models	4333
<i>Mengzhao Wang, Haotian Wu, Xiangyu Ke, Yunjun Gao, Xiaoliang Xu, Lu Chen</i>	
DBG-PT: A Large Language Model Assisted Query Performance Regression Debugger	4337
<i>Victor Giannakouris, Immanuel Trummer</i>	
Rodeo: Making Refinements for Diverse Top-k Queries	4341
<i>Felix S. Campbell, Julia Stoyanovich, Yuval Moskovitch</i>	

QPJVis Demo: Quality-boost Progressive Join Query Processing System.....	4345
<i>Xin Zhang, Ahmed Eldawy</i>	
Counterfactual Explanation Analytics: Empowering Lay Users to Take Action Against Consequential Automated Decisions	4349
<i>Peter M. Vannostrand, Dennis M. Hofmann, Lei Ma, Belisha Genin, Randy Huang, Elke A. Rundensteiner</i>	
UniView: A Unified Autonomous Materialized View Management System for Various Databases.....	4353
<i>Zhenrong Xu, Pengfei Wang, Guoze Xue, Qitong Yan, Shenghao Gong, Yelan Jiang, Yuren Mao, Yunjun Gao, Shu Shen, Wei Zhang, Dan Luo, Lu Chen</i>	
A Demonstration of TENDS: Time Series Management System based on Model Selection	4357
<i>Yuanyuan Yao, Shenjia Dai, Yilin Li, Lu Chen, Dimeng Li, Yunjun Gao, Tianyi Li</i>	
SEER: An End-to-End Toolkit for Benchmarking Time Series Database Systems in Monitoring Applications	4361
<i>Luca Althaus, Mourad Khayati, Abdelouahab Khelifati, Anton Dignös, Djellel Difallah, Philippe Cudre-Mauroux</i>	
Demonstration of DB-GPT: Next Generation Data Interaction System Empowered by Large Language Models	4365
<i>Siqiao Xue, Danrui Qi, Caigao Jiang, Fangyin Cheng, Keting Chen, Zhiping Zhang, Hongyang Zhang, Ganglin Wei, Wang Zhao, Fan Zhou, Hong Yi, Shaodong Liu, Hongjun Yang, Faqiang Chen</i>	
DeepSketch: A Query Sketching Interface for Deep Time Series Similarity Search	4369
<i>Zheng Zhang, Joey Shao, Andrew Crotty</i>	
Rock: Cleaning Data with both ML and Logic Rules	4373
<i>Zian Bao, Binbin Bie, Wenfei Fan, Daji Li, Mengyun Li, Kaiwen Lin, Wei Lin, Peijie Liu, Peng Liu, Zhicong Lv, Mingliang Ouyang, Chenyang Sun, Shuai Tang, Yaoshu Wang, Qiyuan Wei, Xiangqian Wu, Min Xie, Jing Zhang, Zhao Runxiao, Jie Zhu, Yilin Zhu</i>	
Clean4TSDB: A Data Cleaning Tool for Time Series Databases.....	4377
<i>Xiaou Ding, Song Yichen, Hongzhi Wang, Donghua Yang, Chen Wang, Jianmin Wang</i>	
LakeCompass: An End-to-End System for Table Maintenance, Search and Analysis in Data Lakes	4381
<i>Chengliang Chai, Yuhao Deng, Yutong Zhan, Ziqi Cao, Yuanfang Zhang, Lei Cao, Yu-Ping Wang, Zhiwei Zhang, Ye Yuan, Guoren Wang, Nan Tang</i>	
DOP-SQL: A General-purpose, High-utility, and Extensible Private SQL System	4385
<i>Jianzhe Yu, Wei Dong, Juanru Fang, Dajun Sun, Ke Yi</i>	
Catcher: A Cache Analysis System for Top-k Pub/Sub Service	4389
<i>Baolong Mei, Yafei Li, Wei Chen, Linshen Luan, Guanglei Zhu, Yuanyuan Jin, Jianliang Xu</i>	
Optimizing Distributed Tiered Data Storage Systems with DITIS	4393
<i>Sotiris Vasileiadis, Matthew Paraskeva, George Savva, Andreas Efstathiou, Edson Ramiro Lucas Filho, Jianqiang Shen, Lun Yang, Kebo Fu, Herodotos Herodotou</i>	
VQFT: A Visual Query Approach Based on Full-Text Search for Knowledge Graphs	4397
<i>Zhaozhuo Li, Xin Wang, Meng Wang, Yajun Yang, Bohan Li, Dong Han</i>	
CORAL: Collaborative Automatic Labeling System based on Large Language Models	4401
<i>Zhen Zhu, Yibo Wang, Shouqing Yang, Lin Long, Runze Wu, Xiu Tang, Junbo Zhao, Haobo Wang</i>	

CMixing: An Efficient Coin Mixing Platform to Enhance Anonymity in Cryptocurrency Transactions	4405
<i>Wangze Ni, Yiwei Zhao, Pengze Chen, Lei Chen, Peng Cheng, Chen Zhang</i>	
CyNetDiff: A Python Library for Accelerated Implementation of Network Diffusion Models	4409
<i>Eilott W. Robson, Dharmath Reddy, Abhishek Kumar Umrawal</i>	
EncChain: Enhancing Large Language Model Applications with Advanced Privacy Preservation Techniques	4413
<i>Zhe Fu, Mo Sha, Yiran Li, Huorong Li, Yubing Ma, Sheng Wang, Feifei Li</i>	
FairEM360: A Suite for Responsible Entity Matching	4417
<i>Nima Shahbazi, Mahdi Erfanian, Abolfazl Asudeh, Fatemeh Nargesian, Divesh Srivastava</i>	
RetClean: Retrieval-Based Tabular Data Cleaning Using LLMs and Data Lakes.....	4421
<i>Zan Ahmad Naeem, Mohammad Shahmeer Ahmad, Mohamed Eltabakh, Mourad Ouzzani, Nan Tang</i>	
Mach: Firefighting Time-Critical Issues in Complex Systems Using High-Frequency Telemetry.....	4425
<i>Franco Solleza, Shihang Li, William Sun, Richard Tang, Malte Schwarzkopf, Nesime Tatbul, Andrew Crotty, David Cohen, Stan Zdonik</i>	
SketchQL Demonstration: Zero-shot Video Moment Querying with Sketches	4429
<i>Renzhi Wu, Pramod Chunduri, Dristi Shah, Ashmitha Julius Aravind, Ali Payani, Xu Chu, Joy Arulraj, Kexin Rong</i>	
DataPrice: An Interactive System for Pricing Datasets in Data Marketplaces.....	4433
<i>Zhu Yiding, Hongwei Zhang, Jiayao Zhang, Jinfei Liu, Kui Ren</i>	
Demonstration of the VeriEQL Equivalence Checker for Complex SQL Queries	4437
<i>Pinhan Zhao, Yang He, Xinyu Wang, Yuepeng Wang</i>	
FedSQ: A Secure System for Federated Vector Similarity Queries	4441
<i>Zeqi Zhu, Zeheng Fan, Yuxiang Zeng, Yexuan Shi, Yi Xu, Mengmeng Zhou, Jin Dong</i>	
FedSM: A Practical Federated Shared Mobility System	4445
<i>Shuyue Wei, Yuanyuan Zhang, Zimu Zhou, Tianlong Zhang, Ke Xu</i>	
DataLoom: Simplifying Data Loading with LLMs	4449
<i>Alexander Van Renen, Mihail Stoian, Andreas Kipf</i>	
Demonstration of VCR: A Tabular Data Slicing Approach to Understanding Object Detection Model Performance.....	4453
<i>Jie Jeff Xu, Saahir Dhanani, Jorge Piazzentin Ono, Wenbin He, Liu Ren, Kexin Rong</i>	
ModsNet: Performance-aware Top-k Model Search using Exemplar Datasets	4457
<i>Mengying Wang, Hanchao Ma, Sheng Guan, Yiyang Bian, Haolai Che, Abhishek Daundkar, Alp Sehirlioglu, Yinghui Wu</i>	
OFL-W3: A One-shot Federated Learning System on Web 3.0.....	4461
<i>Linshan Jiang, Moming Duan, Bingsheng He, Yulin Sun, Peishen Yan, Yang Hua, Tao Song</i>	
Swift: A Data-Driven Flight Planning System at Scale	4465
<i>Chang Gao, Tianlong Zhang, Yuxiang Zeng, Yi Xu, Shuyuan Li, Yuanyuan Zhang</i>	

Pyneapple-G: Scalable Spatial Grouping Queries	4469
<i>Laila Abdelhafeez, Andres Calderon, Amr Magdy, Vassilis J. Tsotras</i>	
PD-Explain: A Unified Python-native Framework for Query Explanations Over DataFrames	4473
<i>Itay Elyashiv, Amir Gilad, Edna Isakov, Tal Tikochinsky, Amit Somech</i>	
HocoPG: A Database System with Homomorphic Compression for Text Processing	4477
<i>Jiawei Guan, Feng Zhang, Yuxin Tang, Weitang Ye, Xiaoyong Du</i>	
Chat2Data: An Interactive Data Analysis System with RAG, Vector Databases and LLMs	4481
<i>Xinyang Zhao, Xuanhe Zhou, Guoliang Li</i>	
PrismX: A Single-Machine System for Querying Big Graphs	4485
<i>Shuhao Liu, Yang Liu, Wenfei Fan</i>	
TimeCSL: Unsupervised Contrastive Learning of General Shapelets for Explorable Time Series Analysis	4489
<i>Zhiyu Liang, Chen Liang, Zheng Liang, Hongzhi Wang, Bo Zheng</i>	
HSAP: A Human-in-the-loop Social Media-based Situation Awareness Platform	4493
<i>Xiangmin Zhou, Chengkun He, Xi Chen, Yanchun Zhang</i>	
DB-MAGS: Multi-Anomaly Data Generation System for Transactional Databases.....	4497
<i>Yiqi Shen, Sijia Li, Miaodong Shen, Peng Cai, Weiyuan Xu, Li Kai, Jinlong Cai</i>	
QuoteInspector: Gaining Insight about Social Media Discussions	4501
<i>Peizhi Wu, Yi Zhang, Wang-Chiew Tan, Zack G. Ives</i>	

Panel

Vector Databases: What's Really New and What's Next?	4505
<i>Jianguo Wang, Eric Hanson, Guoliang Li, Yannis Papakonstantinou, Harsha Simhadri, Charles Xie</i>	

Endowment Awards

High-Performance Spatial Data Analytics: Systematic R&D for Scale-Out and Scale-Up Solutions from the Past to Now	4507
<i>Fusheng Wang, Rubao Lee, Dejun Teng, Xiaodong Zhang, Joel Saltz</i>	
Intelligent Agents for Data Exploration.....	4521
<i>Siheh Amer-Yahia</i>	
Reimagining Deep Learning Systems Through the Lens of Data Systems	4531
<i>Arun Kumar</i>	
LeanStore: A High-Performance Storage Engine for NVMe SSDs	4536
<i>Viktor Leis</i>	

Keynotes

Databases Unbound: Querying All of the World's Bytes with AI	4546
<i>Samuel Madden, Michael Cafarella, Michael Franklin, Tim Kraska</i>	

Sharing Information with Differential Privacy: A Database Perspective.....	4555
<i>Xiaokui Xiao</i>	
Harmonizing ML and Databases: A Symphony of Data	4556
<i>Fatma Ozcan</i>	