

51st International Conference on Very Large Data Bases (VLDB 2025)

Proceedings of the VLDB Endowment Volume 18

London, United Kingdom
1-5 September 2025

Part 1 of 8

ISBN: 979-8-3313-3233-4

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. (2026)

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

The Key to Effective UDF Optimization: Before Inlining, First Perform Outlining.....	1
<i>Samuel Arch, Yuchen Liu, Todd Mowry, Jignesh Patel, Andrew Pavlo</i>	
CUTTANA: Scalable Graph Partitioning for Faster Distributed Graph Databases and Analytics.....	14
<i>Milad Rezaei Hajidehi, Sraavan Sridhar, Margo Seltzer</i>	
Cardinality Estimation for Having-Clauses	28
<i>Guido Moerkotte</i>	
Chameleon: a Heterogeneous and Disaggregated Accelerator System for Retrieval-Augmented Language Models	42
<i>Wenqi Jiang, Marco Zeller, Roger Waleffe, Torsten Hoefler, Gustavo Alonso</i>	
LLM-R2: A Large Language Model Enhanced Rule-based Rewrite System for Boosting Query Efficiency.....	53
<i>Zhaodonghui Li, Haitao Yuan, Huiming Wang, Gao Cong, Lidong Bing</i>	
Nitro: Boosting Distributed Reinforcement Learning with Serverless Computing	66
<i>Hanfei Yu, Jacob Carter, Hao Wang, Devesh Tiwari, Jian Li, Seung-Jong Park</i>	

TABLE OF CONTENTS

Research Papers

RED: Effective Trajectory Representation Learning with Comprehensive Information	80
<i>Silin Zhou, Shuo Shang, Lisi Chen, Christian S. Jensen, Panos Kalnis</i>	
Accurate and Fast Approximate Graph Pattern Mining at Scale.....	93
<i>Anna A Arpaci-Dusseau, Zixiang Zhou, Xu hao Chen</i>	
QueryArtisan: Generating Data Manipulation Codes for Ad-hoc Analysis in Data Lakes.....	108
<i>Xiu Tang, Wenhao Liu, Sai Wu, Chang Yao, Gongsheng Yuan, Shanshan Ying, Gang Chen</i>	
Efficient and Effective Algorithms for A Family of Influence Maximization Problems with A Matroid Constraint.....	117
<i>Yiqian Huang, Shiqi Zhang, Laks V.S. Lakshmanan, Wenqing Lin, Xiaokui Xiao, Bo Tang</i>	
COLOR: A Framework for Applying Graph Coloring to Subgraph Cardinality Estimation	130
<i>Kyle B Deeds, Diandre Miguel B Sabale, Moe Kayali, Dan Suciu</i>	
Fully Automated Correlated Time Series Forecasting in Minutes	144
<i>Xinle Wu, Xingjian Wu, Dalin Zhang, Miao Zhang, Chenjuan Guo, Bin Yang, Christian S. Jensen</i>	
From Logs to Causal Inference: Diagnosing Large Systems	158
<i>Markos Markakis, Brit Youngmann, Trinity Gao, Ziyu Zhang, Rana Shahout, Peter Baile Chen, Chunwei Liu, Ibrahim Sabek, Michael Cafarella</i>	
NeutronTP: Load-Balanced Distributed Full-Graph GNN Training with Tensor Parallelism.....	173
<i>Xin Ai, Hao Yuan, Zeyu Ling, Qiange Wang, Yanfeng Zhang, Zhenbo Fu, Chaoyi Chen, Yu Gu, Ge Yu</i>	
Accuracy-enhanced Sparse Vector Technique with Exponential Noise and Optimal Threshold Correction	187
<i>Yuhan Liu, Sheng Wang, Yixuan Liu, Feifei Li, Hong Chen</i>	
Maximum Defective Clique Computation: Improved Time Complexities and Practical Performance	200
<i>Lijun Chang</i>	
Substructure-aware Log Anomaly Detection	213
<i>Yanni Tang, Zhuoxing Zhang, Kaiqi Zhao, Lanting Fang, Zhenhua Li, Wu Chen</i>	
Less is More: Efficient Time Series Dataset Condensation via Two-fold Modal Matching	226
<i>Hao Miao, Ziqiao Liu, Yan Zhao, Chenjuan Guo, Bin Yang, Kai Zheng, Christian S. Jensen</i>	
A Memory Guided Transformer for Time Series Forecasting.....	239
<i>Yunyao Cheng, Chenjuan Guo, Bin Yang, Haomin Yu, Kai Zhao, Christian S. Jensen</i>	

LEAP: LLM-powered End-to-end Automatic Library for Processing Social Science Queries on Unstructured Data.....	253
<i>Chuxuan Hu, Austin Peters, Daniel Kang</i>	
TEAM: Topological Evolution-aware Framework for Traffic Forecasting.....	265
<i>Duc Kieu, Tung Kieu, Peng Han, Bin Yang, Christian S. Jensen, Bac Le</i>	
Chimera: A system design of dual storage and traversal-join unified query processing for SQL/PGQ	279
<i>Geonho Lee, Jeongho Park, Min-Soo Kim</i>	
Can Graph Reordering Speed Up Graph Neural Network Training? An Experimental Study	293
<i>Nikolai Merkel, Pierre Toussing, Ruben Mayer, Hans-Arno Jacobsen</i>	
HyperBlocker: Accelerating Rule-based Blocking in Entity Resolution using GPUs	308
<i>Xiaoke Zhu, Min Xie, Ting Deng, Qi Zhang</i>	
Calibrating Noise for Group Privacy in Subsampled Mechanisms.....	322
<i>Yangfan Jiang, Xinjian Luo, Yin Yang, Xiaokui Xiao</i>	
Outback: Fast and Communication-efficient Index for Key-Value Store on Disaggregated Memory	335
<i>Yi Liu, Minghao Xie, Shouqian Shi, Yuanchao Xu, Heiner Litz, Chen Qian</i>	
Making CRDTs Not So Eventual.....	349
<i>Yunhao Mao, Gengrui Zhang, Zongxin Liu, Pezhman Nasirifard, Sofia Tijanac, Hans-Arno Jacobsen</i>	
Maximum k-Plex Search: An Alternated Reduction-and-Bound Method	363
<i>Shuohao Gao, Kaiqiang Yu, Shengxin Liu, Cheng Long</i>	
Discovering Leitmotifs in Multidimensional Time Series.....	377
<i>Patrick Schäfer, Ulf Leser</i>	
SIMformer: Single-Layer Vanilla Transformer Can Learn Free-Space Trajectory Similarity	390
<i>Chuang Yang, Renhe Jiang, Xiaohang Xu, Chuan Xiao, Kaoru Sezaki</i>	
CUBIT: Concurrent Updatable Bitmap Indexing.....	399
<i>Junchang Wang, Manos Athanassoulis</i>	
Privacy-Enhanced Database Synthesis for Benchmark Publishing.....	413
<i>Yongrui Zhong, Yunqing Ge, Jianbin Qin, Shuyuan Zheng, Bo Tang, Yu-Xuan Qiu, Rui Mao, Ye Yuan, Makoto Onizuka, Chuan Xiao</i>	
Themis: A GPU-accelerated Relational Query Execution Engine	426
<i>Kijae Hong, Kyoungmin Kim, Young-Koo Lee, Yang-Sae Moon, Sourav S Bhowmick, Wook-Shin Han</i>	
Finding Convincing Views to Endorse a Claim	439
<i>Shunit Agmon, Amir Gilad, Brit Youngmann, Shahar Zoarets, Benny Kimelfeld</i>	
Quantifying Point Contributions: A Lightweight Framework for Efficient and Effective Query-Driven Trajectory Simplification.....	453
<i>Yumeng Song, Yu Gu, Tianyi Li, Yushuai Li, Christian S. Jensen, Ge Yu</i>	
MILLION: A General Multi-Objective Framework with Controllable Risk for Portfolio Management	466
<i>Liwei Deng, Tianfu Wang, Yan Zhao, Kai Zheng</i>	

The Cost of Representation by Subset Repairs	475
<i>Yuxi Liu, Fangzhu Shen, Kushagra Ghosh, Amir Gilad, Benny Kimelfeld, Sudeepa Roy</i>	
cedar: Optimized and Unified Machine Learning Input Data Pipelines	488
<i>Mark Zhao, Emanuel Adamiak, Christos Kozyrakis</i>	
Goku: A Schemaless Time Series Database for Large Scale Monitoring at Pinterest	503
<i>Monil Mukesh Sanghavi, Ming-May Hu, Zhenxiao Luo, Xiao Li, Kapil Bajaj</i>	

TABLE OF CONTENTS

Research Papers

Agent-OM: Leveraging LLM Agents for Ontology Matching	516
<i>Zhangcheng Qiang, Weiqing Wang, Kerry Taylor</i>	
GraphAr: An Efficient Storage Scheme for Graph Data in Data Lakes.....	530
<i>Xue Li, Weibin Zeng, Zhibin Wang, Diwen Zhu, Jingbo Xu, Wenyuan Yu, Jingren Zhou</i>	
Cardinality Estimation for Similarity Search on High-Dimensional Data Objects: The Impact of Reference Objects	544
<i>Hai Lan, Shixun Huang, Zhifeng Bao, Renata Borovica-Gajic</i>	
Efficient Top-k Frequent Subgraph Mining Using Tight Upper and Lower Bounds	557
<i>Seonho Lee, Yeunjun Lee, Kunsoo Park</i>	
MSGNN: Masked Schema based Graph Neural Networks.....	571
<i>Hao Liu, Qianwen Yang, Taoyong Cui, Wei Wang</i>	
How Reliable Are Streams? End-to-End Processing-Guarantee Validation and Performance Benchmarking of Stream Processing Systems	585
<i>Jawad Tahir, Ruben Mayer, Christoph Doblender, Hans-Arno Jacobsen</i>	
Towards Sufficient GPU-accelerated Dynamic Graph Management: Survey and Experiment	599
<i>Yinnian Lin, Lei Zou, Xunbin Su</i>	
RDPro: Distributed Processing of Big Raster Data	613
<i>Zhuocheng Shang, Samridhi Singla, Ahmed Eldawy, Elia Scudiero</i>	
Approximate Anchored Densest Subgraph Search on Large Static and Dynamic Graphs	623
<i>Qi Zhang, Yalong Zhang, Ronghua Li, Guoren Wang</i>	
PRICE: A Pretrained Model for Cross-Database Cardinality Estimation.....	637
<i>Tianjing Zeng, Junwei Lan, Jiahong Ma, Wenqing Wei, Rong Zhu, Pengfei Li, Bolin Ding, Defu Lian, Zhewei Wei, Jingren Zhou</i>	
Datalog with First-Class Facts.....	651
<i>Thomas Gilray, Arash Sahebolamri, Yihao Sun, Sowmith Kunapaneni, Sidharth Kumar, Kristopher Micinski</i>	
T-Assess: An Efficient Data Quality Assessment System Tailored for Trajectory Data.....	666
<i>Junhao Zhu, Tao Wang, Danlei Hu, Ziquan Fang, Lu Chen, Yunjun Gao, Tianyi Li, Christian S. Jensen</i>	
LEAP: A Low-cost Spark SQL Query Optimizer using Pairwise Comparison.....	675
<i>Junhao Ye, Jiahui Li, Lu Chen, Yuren Mao, Yunjun Gao, Tianyi Li</i>	
Towards Practical Oblivious Map.....	688
<i>Xinle Cao, Weiqi Feng, Jian Liu, Jinjin Zhou, Wenjing Fang, Lei Wang, Quanqing Xu, Chuanhui Yang, Kui Ren</i>	

PolyBase: Adapting to Data Affinity Changes in Geo-Replicated Database via Row-Level Paxos-Group Affiliation Re-Assignment.....	702
<i>Chaoyi Ruan, Yingqiang Zhang, Juncheng Zhang, Cheng Li, Xiaosong Ma, Hao Chen, Jie Zhou, Feifei Li, Xinjun Yang</i>	
Explaining GNN-based Recommendations in Logic.....	715
<i>Wenfei Fan, Lihang Fan, Dandan Lin, Min Xie</i>	
Efficient Computation of Hyper-triangles on Hypergraphs.....	729
<i>Haozhe Yin, Kai Wang, Wenjie Zhang, Ying Zhang, Ruijia Wu, Xuemin Lin</i>	
Laser: Buffer-Aware Learned Query Scheduling in Master-Standby Databases.....	743
<i>Yuwei Huang, Guoliang Li</i>	
A Single Machine System for Querying Big Graphs with PRAM.....	756
<i>Yang Liu, Wenfei Fan, Shuhao Liu, Xiaohe Zhu, Jianxin Li</i>	
A CPU-GPU Hybrid Labelling Algorithm for Massive Shortest Distance Queries on Road Networks.....	770
<i>Li Jiajia, Yongzhi Chen, Mengxuan Zhang, Lei Li</i>	
SDEcho: Efficient Explanation of Aggregated Sequence Difference.....	784
<i>Fei Ye, Zikang Liu, Xi Zhang, Yinan Jing, Zhenying He, Yuxin Che, Haoran Xiong, Kai Zhang, X. Sean Wang</i>	
MLP-Mixer based Masked Autoencoders Are Effective, Explainable and Robust for Time Series Anomaly Detection.....	798
<i>Tang Qideng, Dai Chaofan, Wu Yahui, Zhou Haohao</i>	
Efficient Data-aware Distance Comparison Operations for High-Dimensional Approximate Nearest Neighbor Search.....	812
<i>Liwei Deng, Penghao Chen, Ximu Zeng, Tianfu Wang, Yan Zhao, Kai Zheng</i>	
Seer: Accelerating Blockchain Transaction Execution by Fine-Grained Branch Prediction.....	822
<i>Shijie Zhang, Ru Cheng, Xinpeng Liu, Jiang Xiao, Hai Jin, Bo Li</i>	
The ParClusterers Benchmark Suite (PCBS): A Fine-Grained Analysis of Scalable Graph Clustering.....	836
<i>Shangdi Yu, Jessica Shi, Jamison Meindl, David Eisenstat, Xiaoen Ju, Sasan Tavakkol, Laxman Dhulipala, Jakub Łącki, Vahab Mirrokni, Julian Shun</i>	
Semantic Conformance Testing of Relational DBMS.....	850
<i>Shuang Liu, Chenglin Tian, Jun Sun, Ruifeng Wang, Wei Lu, Yongxin Zhao, Yinxing Xue, Junjie Wang, Xiaoyong Du</i>	
RankPQO: Learning-to-Rank for Parametric Query Optimization.....	863
<i>Songsong Mo, Yue Zhao, Zhifeng Bao, Quanqing Xu, Chuanhui Yang, Gao Cong</i>	
Datamap-Driven Tabular Coreset Selection for Classifier Training.....	876
<i>Aviv Hadar, Tova Milo, Kathy Razmadze</i>	
Towards Scalable and Practical Batch-Dynamic Connectivity.....	889
<i>Quinten De Man, Laxman Dhulipala, Adam Karczmarz, Jakub Łącki, Julian Shun, Zhongqi Wang</i>	
IcedTea: Efficient and Responsive Time-Travel Debugging in Dataflow Systems.....	902
<i>Shengquan Ni, Yicong Huang, Zuozhi Wang, Chen Li</i>	

Representative Time Series Discovery for Data Exploration..... 915
Ge Lee, Shixun Huang, Zhifeng Bao, Yanchang Zhao

TABLE OF CONTENTS

Research Papers

Efficient Graph Embedding Generation and Update for Large-Scale Temporal Graph.....	929
<i>Yifan Song, Xiaolong Chen, Wenqing Lin, Jia Li, Chen Zhang, Yan Zhou, Lei Chen, Jing Tang</i>	
FaDE: More Than a Million What-ifs Per Second.....	943
<i>Haneen Mohammed, Eugene Wu, Alexander Yao, Charlie Summers, Lampros Flokas, Gromit Yeuk-Yin Chan, Subrata Mitra, Hongbin Zhong</i>	
Towards Ideal Temporal Graph Neural Networks: Evaluations and Conclusions after 10,000 GPU Hours	956
<i>Yuxin Yang, Hongkuan Zhou, Rajgopal Kannan, Viktor Prasanna</i>	
Kishu: Time-Traveling for Computational Notebooks.....	970
<i>Zhaoheng Li, Supawit Chockchawat, Areet Sheth, Yongjoo Park, Ribhav Sahu</i>	
GTI: Graph-based Tree Index with Logarithm Updates for Nearest Neighbor Search in High-Dimensional Spaces	986
<i>Ruiyao Ma, Yifan Zhu, Baihua Zheng, Lu Chen, Congcong Ge, Yunjun Gao</i>	
Incremental Detection of Denial Constraint Violations.....	1000
<i>Youri Kaminsky, Eduardo Pena, Felix Naumann</i>	
Revisiting CNNs for Trajectory Similarity Learning.....	1013
<i>Zhihao Chang, Linzhu Yu, Huan Li, Sai Wu, Gang Chen, Dongxiang Zhang</i>	
Most Similar Biclique Search at Scale	1022
<i>Deming Chu, Zhizhi Gao, Fan Zhang, Wenjie Zhang, Xuemin Lin, Zhihong Tian</i>	
SPECIAL: Synopsis Assisted Secure Collaborative Analytics	1035
<i>Chenghong Wang, Lina Qiu, Johes Bater, Yukui Luo</i>	
IncrCP: Decomposing and Orchestrating Incremental Checkpoints for Effective Recommendation Model Training.....	1049
<i>Qingyin Lin, Jiansu Du, Rui Li, Zhiguang Chen, Wenguang Chen, Nong Xiao</i>	
WeShap: Weak Supervision Source Evaluation with Shapley Values	1063
<i>Naiqing Guan, Nick Koudas</i>	
Are Joins over LSM-trees Ready: Take RocksDB as an Example	1077
<i>Weiping Yu, Fan Wang, Xuwei Zhang, Siqiang Luo</i>	
Interactive Graph Search for Multiple Targets on DAGs.....	1091
<i>Zheng Wu, Xuliang Zhu, Yixiang Fang, Jianliang Xu, Xin Huang</i>	

AdaNDV: Adaptive Number of Distinct Value Estimation via Learning to Select and Fuse Estimators.....	1104
<i>Xianghong Xu, Tieying Zhang, Xiao He, Haoyang Li, Rong Kang, Wang Shuai, Xu Linhui, Zhimin Liang, Shangyu Luo, Lei Zhang, Jianjun Chen</i>	
UNIFY: Unified Index for Range Filtered Approximate Nearest Neighbors Search	1118
<i>Anqi Liang, Pengcheng Zhang, Bin Yao, Zhongpu Chen, Yitong Song, Guangxu Cheng</i>	
In-depth Analysis of Densest Subgraph Discovery in a Unified Framework	1131
<i>Yingli Zhou, Qingshuo Guo, Yi Yang, Yixiang Fang, Chenhao Ma, Laks Lakshmanan</i>	
Sphinteract: Resolving Ambiguities in NL2SQL Through User Interaction.....	1145
<i>Fuheng Zhao, Shaleen Deep, Fotis Psallidas, Avrilia Floratou, Divy Agrawal, Amr El Abbadi</i>	
Noise Matters: Cross Contrastive Learning for Flink Anomaly Detection.....	1159
<i>Zhihao Zhuang, Yingying Zhang, Kai Zhao, Chenjuan Guo, Bin Yang, Qingsong Wen, Lunting Fan</i>	
RCRank: Multimodal Ranking of Root Causes of Slow Queries in Cloud Database Systems	1169
<i>Yang Liu, Wenfei Fan, Shuhao Liu, Xiaoke Zhu, Jianxin Li</i>	
Ranking Indicator Discovery from Very Large Knowledge Graphs	1183
<i>Hassan Abdallah, Béatrice Markhoff, Arnaud Soulet</i>	
Graph Neural Network Training Systems: A Performance Comparison of Full-Graph and Mini-Batch....	1196
<i>Saurabh Bajaj, Hui Guan, Marco Serafini, Juelin Liu, Hojae Son</i>	
Discovering Approximate Inclusion Dependencies	1210
<i>Qingdong Su, Zhikang Wang, Zijing Tan, Shuai Ma</i>	
DumpKV: Learning based lifetime aware garbage collection for key value separation in LSM-tree.....	1223
<i>Zhutao Zhuang, Zhiguang Chen, Xinqi Zeng</i>	
RGS-Sketch: An Accurate, Invertible, and Mergeable Sketch for Online Super Spreader Detection in High-speed Data Streams	1237
<i>Boyu Zhang, He Huang, Yu-E Sun, Guoju Gao</i>	
Vortex: Overcoming Memory Capacity Limitations in GPU-Accelerated Large-Scale Data Analytics	1250
<i>Yichao Yuan, Advait Iyer, Lin Ma, Nishil Talati</i>	

TABLE OF CONTENTS

Research Papers

Dandelion: Smaller Clusters, Bigger Speeds—Distributed Transactions Redefined	1264
<i>Antonios Katsarakis, Vasilis Gavrielatos, Chris Jensen, Nikos Ntarmos</i>	
Esc: An Early-Stopping Checker for Budget-aware Index Tuning.....	1278
<i>Xiaoying Wang, Wentao Wu, Vivek Narasayya, Surajit Chaudhuri</i>	
Jodes: Efficient Oblivious Join in the Distributed Setting	1291
<i>Yilei Wang, Xiangdong Zeng, Sheng Wang, Feifei Li</i>	
OpenFGL: A Comprehensive Benchmark for Federated Graph Learning.....	1305
<i>Xunkai Li, Yinlin Zhu, Boyang Pang, Guochen Yan, Yeyu Yan, Zening Li, Zhengyu Wu, Wentao Zhang, Ronghua Li, Guoren Wang</i>	
Evaluating Continuous Queries with Inconsistency Annotations	1321
<i>Samuele Langhi, Angela Bonifati, Riccardo Tommasini</i>	
On More Efficiently and Versatilely Querying Historical k-Cores	1335
<i>Zhi Wang, Ming Zhong, Yuanyuan Zhu, Tiejun Qian, Mengchi Liu, Jeffrey Xu Yu</i>	
GeoBloom: Revisiting Lightweight Models for Geographic Information Retrieval.....	1348
<i>Yi Li, Gao Cong</i>	
VerIso: Verifiable Isolation Guarantees for Database Transactions.....	1362
<i>Shabnam Ghasemirad, Si Liu, Christoph Sprenger, Luca Multazzu, David Basin</i>	
A Hybrid Approach to Integrating Deterministic and Non-deterministic Concurrency Control in Database Systems	1376
<i>Yinhao Hong, Hongyao Zhao, Wei Lu, Xiaoyong Du, Yuxing Chen, Anqun Pan, Zheng Lixiong</i>	
From Genesis to Maturity: Managing Knowledge Graph Ecosystems Through Life Cycles.....	1390
<i>Sandra Geisler, Cinzia Cappiello, Irene Celino, David Chaves-Fraga, Anastasia Dimou, Ana Iglesias-Molina, Maurizio Lenzerini, Anisa Rula, Dylan Van Assche, Sascha Welten, Maria-Esther Vidal</i>	
Avoiding Materialisation for Guarded Aggregate Queries	1398
<i>Matthias Lanzinger, Reinhard Pichler, Alexander Selzer</i>	
Synergetic Community Search over Large Multilayer Graphs	1412
<i>Chengyang Luo, Qing Liu, Yunjun Gao, Jianliang Xu</i>	
Searching and Detecting Structurally Similar Communities in Large Heterogeneous Information Networks	1425
<i>Shu Wang, Yixiang Fang, Wensheng Luo</i>	
Cabinet: Dynamically Weighted Consensus Made Fast	1439
<i>Gengrui Zhang, Shiquan Zhang, Michail Bachras, Yuqiu Zhang, Hans-Arno Jacobsen</i>	

Mining the Minoria: Unknown, Under-represented, and Under-performing Minority Groups.....	1453
<i>Mohsen Dehghankar, Abolfazl Asudeh</i>	
Efficient Concurrent Updates to Persistent Randomized Binary Search Trees	1481
<i>Guanhao Hou, Jinchao Huang, Fangyuan Zhang, Sibow Wang</i>	
Explaining Black-Box Clustering Pipelines With Cluster-Explorer.....	1495
<i>Sariel Ofek, Amit Somech</i>	
BACH: Bridging Adjacency List and CSR Format using LSM-Trees for HGTAP Workloads	1509
<i>Jianfeng Huang, Cao Yihao, Ren Shubing, Baohua Wu, Dongjing Miao</i>	
FLEET: High-Performance Durable Replicated State Machines using Scattered and Coordinated Log Entries	1522
<i>Hua Fan, Hao Tan, Wenchao Zhou, Feifei Li</i>	
BigVectorBench: Heterogeneous Data Embedding and Compound Queries are Essential in Evaluating Vector Databases.....	1536
<i>Guoxin Kang, Zhongxin Ge, Jingpei Hu, Xueya Zhang, Lei Wang, Jianfeng Zhan</i>	

TABLE OF CONTENTS

Research Papers

A Systematic Study on Early Stopping Metrics in HPO and the Implications of Uncertainty.....	1551
<i>Jiawei Guan, Feng Zhang, Jiesong Liu, Xiaoyong Du, Xipeng Shen</i>	
TELESAFE - Detecting Private/Work Boundary Crossings in Energy Consumption Trails in Telework..	1565
<i>Haoying Zhang, Mariem Brahem, Nicolas Ancaux, Benjamin Nguyen, Jose Maria De Fuentes</i>	
FB+-tree: A Memory-Optimized B+-tree with Latch-Free Update.....	1579
<i>Yuan Chen, Ao Li, Wenhai Li, Lingfeng Deng</i>	
VStream: A Distributed Streaming Vector Search System	1593
<i>Shenghao Gong, Haobo Sun, Ziquan Fang, Liu Liu, Lu Chen, Yunjun Gao</i>	
Efficient Historical Butterfly Counting in Large Temporal Bipartite Networks via Graph Structure-aware Index.....	1607
<i>Qiuyang Mang, Jingbang Chen, Hangrui Zhou, Yu Gao, Yingli Zhou, Qingyu Shi, Richard Peng, Yixiang Fang, Chenhao Ma</i>	
PlanRGCN: Predicting SPARQL Query Performance	1621
<i>Abiram Mohanaraj, Matteo Lissandrini, Katja Hose</i>	
Holistic query Approximation via RL Modeling.....	1635
<i>Susan Davidson, Tova Milo, Kathy Razmadze, Gal Zeevi</i>	
Unleashing Graph Partitioning for Large-Scale Nearest Neighbor Search.....	1649
<i>Lars Gottesbueren, Laxman Dhulipala, Rajesh Jayaram, Jakub Łacki</i>	
BiST: A Lightweight and Efficient Bi-directional Model for Spatiotemporal Prediction	1663
<i>Jiaming Ma, Binwu Wang, Pengkun Wang, Zhengyang Zhou, Xu Wang, Yang Wang</i>	
QOVIS: Understanding and Diagnosing Query Optimizer via a Visualization-assisted Approach.....	1677
<i>Zhengxin You, Qiaomu Shen, Man Lung Yiu, Bo Tang</i>	
Unsupervised Anomaly Detection in Multivariate Time Series across Heterogeneous Domains	1691
<i>Vincent Jacob, Yanlei Diao</i>	
NeutronTask: Scalable and Efficient Multi-GPU GNN Training with Task Parallelism	1705
<i>Zhenbo Fu, Xin Ai, Qiange Wang, Yanfeng Zhang, Shizhan Lu, Chaoyi Chen, Chunyu Cao, Hao Yuan, Zhewei Wei, Yu Gu, Yingyou Wen, Ge Yu</i>	
Quantum Data Management in the NISQ Era	1720
<i>Rihan Hai, Shih-Han Hung, Tim Coopmans, Tim Littau, Floris Geerts</i>	
G-View: View Management for Graph Databases.....	1730
<i>Yunjia Zheng, Charlotte Sacré, Mohanna Shahrads, Owen Lipchitz, Yuting Gu, Bettina Kemme</i>	

Privacy for Free: Leveraging Local Differential Privacy Perturbed Data from Multiple Services.....	1743
<i>Rong Du, Qingqing Ye, Yue Fu, Haibo Hu</i>	
K2: On Optimizing Distributed Transactions in a Multi-region Data Store with True-time Clocks.....	1756
<i>Haoze Song, Yongqi Wang, Xusheng Chen, Hao Feng, Yazhi Feng, Xieyun Fang, Heming Cui, Linghe Kong</i>	
Maximum Inner Product is Query-Scaled Nearest Neighbor.....	1770
<i>Tingyang Chen, Cong Fu, Kun Wang, Xiangyu Ke, Yunjun Gao, Wenchao Zhou, Yabo Ni, Anxiang Zeng</i>	
Migration-Free Elastic Storage of Time Series in Apache IoTDB.....	1784
<i>Rongzhao Chen, Xiangpeng Hu, Xiangdong Huang, Chen Wang, Shaoxu Song, Jianmin Wang</i>	
GQL and SQL/PGQ: Theoretical Models and Expressive Power	1798
<i>Amelie Gheerbrant, Leonid Libkin, Liat Peterfreund, Alexandra Rogova</i>	
A Practical Theory of Generalization in Selectivity Learning	1811
<i>Peizhi Wu, Haoshu Xu, Ryan Marcus, Zack Ives</i>	
Revisiting the Index Construction of Proximity Graph-Based Approximate Nearest Neighbor Search....	1825
<i>Shuo Yang, Jiadong Xie, Yingfan Liu, Jeffrey Xu Yu, Xiyue Gao, Qianru Wang, Yanguo Peng, Jiangtao Cui</i>	
Mining Platoon Patterns from Traffic Videos.....	1839
<i>Yijun Bei, Teng Ma, Dongxiang Zhang, Sai Wu, Kian-Lee Tan, Gang Chen</i>	
Agamoto: Scheduling of Deadline-Oriented Incremental Query Execution under Uncertain Resource Price	1852
<i>Botong Huang, Weng Lianggui, Wei Chen, Zuozhi Wang, Kai Zeng, Chen Li, Yihui Feng, Bolin Ding, Jingren Zhou</i>	
SCompression: Enhancing Database Knob Tuning Efficiency Through Slice-Based OLTP Workload Compression.....	1865
<i>Baoqing Cai, Yu Liu, Lin Ma, Pingqi Huang, Bingcheng Lian, Ke Zhou, Jia Yuan, Jie Yang, Xiaofan Cai, Peijun Wu</i>	
Fucci: Database Transaction Fuzzing via Random Conflict Construction and Multilevel Constraint Solving	1879
<i>Xiyue Gao, Zhuang Liu, Yiran Shen, Hui Li, Yingfan Liu, Hongjun Xiao, Yanguo Peng, Jiangtao Cui</i>	
Streaming Time Series Subsequence Anomaly Detection: A Glance and Focus Approach	1892
<i>Wenjing Wang, Ziyang Yue, Bolong Zheng</i>	
Infinite Stream Estimation under Personalized w-Event Privacy.....	1905
<i>Leilei Du, Peng Cheng, Lei Chen, Heng Tao Shen, Xuemin Lin, Wei Xi</i>	
GPEmu: A GPU Emulator for Faster and Cheaper Prototyping and Evaluation of Deep Learning System Research	1919
<i>Meng Wang, Gus Waldspurger, Naufal Ananda, Yuyang Huang, Kemas Wiharja, John Bent, Swaminathan Sundararaman, Vijay Chidambaram, Haryadi Gunawi</i>	
Causal DAG Summarization	1933
<i>Anna Zeng, Michael Cafarella, Batya Kenig, Markos Markakis, Brit Youngmann, Babak Salimi</i>	
mLoRA: Fine-Tuning LoRA Adapters via Highly-Efficient Pipeline Parallelism in Multiple GPUs.....	1948
<i>Zhengmao Ye, Dengchun Li, Zetao Hu, Tingfeng Lan, Jian Sha, Shicong Zhang, Lei Duan, Jie Zuo, Hui Lu, Yuanchun Zhou, Mingjie Tang</i>	

Anarchy in the Database: A Survey and Evaluation of Database Management System Extensibility	1962
<i>Abigale Kim, Marco Slot, David Andersen, Andrew Pavlo</i>	
A Flexible Framework for Query-oriented Interactive Community Search	1977
<i>Longxu Sun, Xin Huang, Jiannan Wang, Jianliang Xu</i>	
Tabular: Efficiently Building Efficient Indexes	1991
<i>Ziyi Yan, Mohamed Farouk Drira, Tianxun Hu, Tianzheng Wang</i>	

TABLE OF CONTENTS

Research Papers

Efficient Maintenance of 2-Hop Labeling Index on Dynamic Small-World Graphs	2005
Yuanyuan Zeng, Yixiang Fang, Kun Chen, Yangfan Li, Chenhao Ma	
Vive la Différence: Practical Diff Testing of Stateful Applications	2018
Kexin Zhu, Michael Whittaker, Srdjan Petrovic, Robert Grandl, Sanjay Ghemawat	
GREAT: Generalized Reservoir Sampling based Triangle Counting Estimation over Streaming Graphs .	2031
Siyue Wu, Dingming Wu, Sinhong Cheuk, Tsz Nam Chan, Kezhong Lu	
Efficient Discovery of Relaxed Functional Dependencies	2044
Mengran Li, Zijing Tan, Honghui Yang, Shuai Ma	
SimRN: Trajectory Similarity Learning in Road Networks based on Distributed Deep Reinforcement Learning.....	2057
Danlei Hu, Yilin Li, Lu Chen, Ziquan Fang, Yushuai Li, Yunjun Gao, Tianyi Li	
BIRDIE: Natural Language-Driven Table Discovery Using Differentiable Search Index	2070
Yuxiang Guo, Zhonghao Hu, Yuren Mao, Baihua Zheng, Yunjun Gao, Mingwei Zhou	
SkyStore: Cost-Optimized Object Storage Across Regions and Clouds.....	2084
Shu Liu, Xiangxi Mo, Moshik Hershcovitch, Henric Zhang, Audrey Cheng, Guy Girmonsky, Gil Vernik, Michael Factor, Tiemo Bang, Soujanya Ponnappalli, Natacha Crooks, Joseph Gonzalez, Danny Harnik, Ion Stoica	
The Power of Constraints in Natural Language to SQL Translation.....	2097
Tonghui Ren, Chen Ke, Yuankai Fan, Yinan Jing, Zhenying He, Kai Zhang, X. Sean Wang	
ACE: A Cardinality Estimator for Set-Valued Queries	2112
Yufan Sheng, Xin Cao, Kaiqi Zhao, Yixiang Fang, Jianzhong Qi, Wenjie Zhang, Christian S. Jensen	
Accio: Bolt-on Query Federation	2126
Xiaoying Wang, Jiannan Wang, Tianzheng Wang, Yong Zhang	
Falcon: Advancing Asynchronous BFT Consensus for Lower Latency and Enhanced Throughput	2136
Xiaohai Dai, Chaozheng Ding, Wei Li, Jiang Xiao, Bolin Zhang, Chen Yu, Albert Zomaya, Hai Jin	
Scalable Pre-Training of Compact Urban Spatio-Temporal Predictive Models on Large-Scale Multi-Domain Data.....	2149
Jindong Han, Hao Wang, Hui Xiong, Hao Liu	
HADES: Range-Filtered Private Aggregation on Public Data.....	2159
Xiaoyuan Liu, Ni Trieu, Trinabh Gupta, Ishtiyaque Ahmad, Dawn Song	
Optimized Batch Prompting for Cost-effective LLMs.....	2172
Zhaoxuan Ji, Xinlu Wang, Zhaojing Luo, Zhongle Xie, Meihui Zhang	

Truss Decomposition in Hypergraphs	2185
Hongchao Qin, Guang Zeng, Ronghua Li, Longlong Lin, Ye Yuan, Guoren Wang	
Optimal Sharding for Scalable Blockchains with Deconstructed SMR.....	2198
Jianting Zhang, Zhongtang Luo, Raghavendra Ramesh, Aniket Kate	
Auto-Prep: Holistic Prediction of Data Preparation Steps for Self-Service Business Intelligence.....	2212
Eugenie Lai, Yeye He, Surajit Chaudhuri	
Time Series Motif Discovery: A Comprehensive Evaluation	2226
Valerio Guerrini, Thibaut Germain, Charles Truong, Laurent Oudre, Paul Boniol	
EinDecomp: Decomposition of Declaratively-Specified Machine Learning and Numerical Computations for Parallel Execution	2240
Daniel Bourgeois, Zhimin Ding, Dimitrije Jankov, Jiehui Li, Mahmoud Sleem, Yuxin Tang, Jiawen Yao, Xinyu Yao, Chris Jermaine	
Continuous Lifelong Conflict-Aware AGV Routing with Kinematic Constraints	2254
Ruizhong Wu, Mengxuan Zhang, Shuxin Wang, Kin Sun Chan, Yan Nei Law, Lei Li	
Wolverine: Highly Efficient Monotonic Search Path Repair for Graph-based ANN Index Updates.....	2268
Dawei Liu, Bolong Zheng, Ziyang Yue, Fuhao Ruan, Xiaofang Zhou, Christian S. Jensen	
Detecting Schema-Related Logic Bugs in Relational DBMSs via Equivalent Database Construction.....	2281
Jiansen Song, Wensheng Dou, Yingying Zheng, Yu Gao, Ziyu Cui, Wei Wang, Jun Wei	
GraphSparseNet: a Novel Method for Large Scale Traffic Flow Prediction.....	2295
Weiyang Kong, Wu Kaiqi, Sen Zhang, Yubao Liu	

TABLE OF CONTENTS

Research Papers

TMLKD: Few-shot Trajectory Metric Learning via Knowledge Distillation	2308
Danling Lai, Jiajie Xu, Jianfeng Qu, Pingfu Chao, Junhua Fang, Chengfei Liu	
Oze: Decentralized Graph-based Concurrency Control for Long-running Update Transactions.....	2321
Jun Nemoto, Taksahi Kambayashi, Takashi Hoshino, Hideyuki Kawashima	
Hermes: Off-the-Shelf Real-Time Transactional Analytics	2334
Elena Milkai, Xiangyao Yu, Jignesh Patel	
Approximation-First Timeseries Monitoring Query At Scale	2348
Zeying Zhu, Jonathan Chamberlain, Kenny Wu, David Starobinski, Zaoxing Liu	
LogCloud: Fast Search of Compressed Logs on Object Storage	2362
Ziheng Wang, Junyu Wei, Alex Aiken, Guangyan Zhang, Jacob Tarring, Rain Jiang, Chenyu Jiang, Wei Xu	
Weak-to-Strong Prompts with Lightweight-to-Powerful LLMs for High-Accuracy, Low-Cost, and Explainable Data Transformation.....	2371
Changlun Li, Chenyu Yang, Yuyu Luo, Ju Fan, Nan Tang	
ChatTS: Aligning Time Series with LLMs via Synthetic Data for Enhanced Understanding and Reasoning	2385
Zhe Xie, Zeyan Li, Xiao He, Longlong Xu, Xidao Wen, Tieying Zhang, Jianjun Chen, Rui Shi, Dan Pei	
Federated Data Shift Distance Estimation	2399
Graham Cormode, Daniel Ting	
Instance-Optimal Acyclic Join Processing Without Regret: Engineering the Yannakakis Algorithm in Column Stores.....	2413
Liese Bekkers, Frank Neven, Stijn Vansummeren, Remy Wang	
Asymmetric Linearizable Local Reads	2427
Myles Thiessen, Guy Khazma, Sam Toueg, Eyal De Lara	
QPET: A Versatile and Portable Quantity-of-Interest-preservation Framework for Error-Bounded Lossy Compression.....	2440
Jinyang Liu, Pu Jiao, Kai Zhao, Xin Liang, Sheng Di, Franck Cappello	
OpenMEL: Unsupervised Multimodal Entity Linking Using Noise-Free Expanded Queries and Global Coherence	2454
Xinyi Zhu, Yongqi Zhang, Lei Chen	
Unraveling the Impact of Window Semantics: Optimizing Join Order for Efficient Stream Processing....	2468
Ariane Ziehn, Jan Szlang, Steffen Zeuch, Volker Markl	

Deduplicated Sampling On-Demand.....	2482
Luca Zecchini, Vasilis Efthymiou, Felix Naumann, Giovanni Simonini	
The Limits of Graph Samplers for Training Inductive Recommender Systems.....	2496
Theis Jendal, Matteo Lissandrini, Peter Dolog, Katja Hose	
HoliPaxos: Towards More Predictable Performance in State Machine Replication.....	2505
Zhiying Liang, Vahab Jabrayilov, Abutalib Aghayev, Aleksey Charapko	
Data-Agnostic Cardinality Learning from Imperfect Workloads.....	2519
Peizhi Wu, Rong Kang, Tieying Zhang, Jianjun Chen, Ryan Marcus, Zack Ives	
BLAEQ: A Multigrid Index for Spatial Query on Geometry Data	2533
Song Wang, Chen Wang, Jianchun Wang, Shengguo Li, Rui Li, Zhiyong Peng	
Simple Testing Can Expose Most Critical Transaction Bugs: Understanding and Detecting Write-Specific Serializability Violations in Database Systems	2547
Ziyu Cui, Wensheng Dou, Yu Gao, Rui Yang, Yingying Zheng, Jiansen Song, Yuan Feng, Jun Wei	
Efficient and Adaptive Estimation of Local Triadic Coefficients	2561
Ilie Sarpe, Aristides Gionis	
VecCity: A Taxonomy-guided Library for Map Entity Representation Learning [Experiment, Analysis \& Benchmark].....	2575
Wentao Zhang, Jingyuan Wang, Yifan Yang, Leong Hou U	
Evaluating Methods for Efficient Entity Count Estimation.....	2589
Jerin George Mathew, Donatella Firmani, Divesh Srivastava	
Fair Transaction Processing For Multi-Tenant Databases.....	2602
Audrey Cheng, Xiao Shi, Aaron Kabcenell, Jolene Huey, Peter Bailis, Natacha Crooks, Ion Stoica	
LobRA: Multi-tenant Fine-tuning over Heterogeneous Data	2616
Sheng Lin, Fangcheng Fu, Haoyang Li, Hao Ge, Xuanyu Wang, Jiawen Niu, Yaofeng Tu, Bin Cui	
Robust Plan Evaluation based on Approximate Probabilistic Machine Learning	2626
Amin Kamali, Verena Kantere, Calisto Zuzarte, Vincent Corvinelli	
CatDB: Data-catalog-guided, LLM-based Generation of Data-centric ML Pipelines	2639
Saeed Fathollahzadeh, Essam Mansour, Matthias Boehm	
Conformal Prediction for Verifiable Learned Query Optimization.....	2653
Hanwen Liu, Shashank Giridhara, Ibrahim Sabek	
Is Integer Linear Programming All You Need for Deletion Propagation? A Unified and Practical Approach for Generalized Deletion Propagation.....	2667
Neha Makhija, Wolfgang Gatterbauer	
Magneto: Combining Small and Large Language Models for Schema Matching	2681
Yurong Liu, Eduardo Pena, Aecio Santos, Eden Wu, Juliana Freire	
Efficient and Accurate Subgraph Counting: A Bottom-up Flow-learning based Approach	2695
Qiuyu Guo, Jianye Yang, Wenjie Zhang, Hanchen Wang, Ying Zhang, Xuemin Lin	

AQETuner: Reliable Query-level Configuration Tuning for Analytical Query Engines	2709
Lixiang Chen, Yuxing Han, Yu Chen, Xing Chen, Chengcheng Yang, Weining Qian	
PipeTGL: (Near) Zero Bubble Memory-based Temporal Graph Neural Network Training via Pipeline Optimization	2722
Jun Liu, Bingqian Du, Ziyue Luo, Sitian Lu, Qiankun Zhang, Hai Jin	
Is Long Context All You Need? Leveraging LLM's Extended Context for NL2SQL	2735
Yeounoh Chung, Gaurav Tarlok Kakkar, Yu Gan, Brenton Milne, Fatma Ozcan	
Saving Private Hash Join	2748
Laurens Kuiper, Paul Gross, Peter Boncz, Hannes Mühleisen	

TABLE OF CONTENTS

Research Papers

Concurrency Control as a Service	2761
Weixing Zhou, Yanfeng Zhang, Xinji Zhou, Zhiyou Wang, Zeshun Peng, Yang Ren, Sihao Li, Huanchen Zhang, Guoliang Li, Ge Yu	
TAB: Unified Benchmarking of Time Series Anomaly Detection Methods.....	2775
Xiangfei Qiu, Zhe Li, Wanghui Qiu, Shiyang Hu, Lekui Zhou, Xingjian Wu, Zhengyu Li, Chenjuan Guo, Aoying Zhou, Zhenli Sheng, Jilin Hu, Christian S. Jensen, Bin Yang	
Heta: Distributed Training of Heterogeneous Graph Neural Networks	2790
Yuchen Zhong, Junwei Su, Chuan Wu, Minjie Wang	
The UDFBench Benchmark for General-purpose UDF Queries	2804
Yannis Foufoulas, Theoni Palaiologou, Alkis Simitsis	
eXpath: Explaining Knowledge Graph Link Prediction with Ontological Closed Path Rules	2818
Ye Sun, Lei Shi, Yongxin Tong	
The LAW theorem: Local Reads and Linearizable Asynchronous Replication	2831
Emmanouil Giortamis, Antonios Katsarakis, Vasilis Gavrielatos, Pramod Bhatotia, Aleksandar Dragojevic, Boris Grot, Vijay Nagarajan, Panagiota Fatourou	
Using Read Promotion and Mixed Isolation Levels for Performant Yet Serializable Execution of Transaction Programs	2846
Brecht Vandervoort, Alan Fekete, Bas Ketsman, Frank Neven, Stijn Vansummeren	
Locality-Aware Cache Replacement Policy for Graph Traversals	2859
Zeynep Korkmaz, Tamer Özsu, Khuzaima Daudjee	
Keigo: Co-designing Log-Structured Merge Key-Value Stores with a Non-Volatile, Concurrency-aware Storage Hierarchy	2872
Rúben Adão, Zhongjie Wu, Changjun Zhou, Oana Balmau, João Paulo, Ricardo Macedo	
Why Are Learned Indexes So Effective but Sometimes Ineffective?	2886
Qiyu Liu, Siyuan Han, Yanlin Qi, Jingshu Peng, Jin Li, Longlong Lin, Lei Chen	
Still More Shades of Null: An Evaluation Suite for Responsible Missing Value Imputation [Experiment, Analysis and Benchmark]	2899
Falaah Arif Khan, Denys Herasymuk, Nazar Protsiv, Julia Stoyanovich	
OpenForge: Probabilistic Metadata Integration.....	2914
Tianji Cong, Fatemeh Nargesian, Junjie Xing, H. V. Jagadish	
Maximum k-Plex Finding: Choices of Pruning Techniques Matter!.....	2928
Akhlaque Ahmad, Da Yan, Xiao Chen, Lyuheng Yuan, Qin Zhang, Saugat Adhikari	

A Comprehensive Survey and Experimental Study of Learning-based Community Search.....	2941
Xiaoxuan Gou, Weiguo Zheng, Yuxiang Wang, Xiaoliang Xu, Zhiyuan Yu	
Decentralized Actor Scheduling and Reference-based Storage in Xorbits: a Native Scalable Data Science Engine	2955
Weizheng Lu, Chao Hui, Yunhai Wang, Feng Zhang, Yueguo Chen, Bao Liu, Chengjie Li, Zhaoxin Wu, Xuye Qin	
STSCache: An Efficient Semantic Caching Scheme for Time-series Data Workloads Based on Hybrid Storage	2964
Tao Kong, Hui Li, Yuxuan Zhao, Liping Li, Xiyue Gao, Qilong Wu, Jiangtao Cui	
Cache Coherence Over Disaggregated Memory	2978
Ruihong Wang, Jianguo Wang, Walid Aref	
Triparts: Scalable Streaming Graph Partitioning to Enhance Community Structure	2992
Ruchi Bhoot, Tuhin Khare, Manoj Agarwal, Siddharth Jaiswal, Yogesh Simmhan	
The LDBC Financial Benchmark: Transaction Workload	3007
Shipeng Qi, Bing Tong, Jiatao Hu, Heng Lin, Yue Pang, Wei Yuan, Songlin Lyu, Zhihui Guo, Ke Huang, Xujin Ba, Qiang Yin, Youren Shen, Yan Zhou, Tao Lv, Jia Li, Lei Zou, Yongwei Wu, Gábor Szárnyas, Xiaowei Zhu, Wenguang Chen, Chuntao Hong	
Alchemy: A Query Optimization Framework for Oblivious SQL	3021
Donghyun Sohn, Kelly Jiang, Nicolas Hammer, Jennie Rogers	
DocETL: Agentic Query Rewriting and Evaluation for Complex Document Processing.....	3035
Shreya Shankar, Tristan Chambers, Tarak Shah, Aditya Parameswaran, Eugene Wu	
HAKES: Scalable Vector Database for Embedding Search Service	3049
Guoyu Hu, Shaofeng Cai, Anh Dinh, Zhongle Xie, Cong Yue, Gang Chen, Beng Chin Ooi	
Path-centric Cardinality Estimation for Subgraph Matching.....	3063
Zhengdong Wang, Qiang Yin, Longbin Lai	
A Comprehensive Study of Shapley Value in Data Analytics.....	3077
Hong Lin, Shixin Wan, Zhongle Xie, Ke Chen, Meihui Zhang, Lidan Shou, Gang Chen	
Effective and Efficient Distributed Temporal Graph Learning through Hotspot Memory Sharing	3093
Longjiao Zhang, Rui Wang, Tongya Zheng, Ziqi Huang, Wenjie Huang, Xinyu Wang, Can Wang, Mingli Song, Sai Wu, Shuibing He	
Stochastic SketchRefine: Scaling In-Database Decision-Making under Uncertainty to Millions of Tuples	3106
Riddho Ridwanul Haque, Anh Mai, Matteo Brucato, Azza Abouzied, Peter Haas, Alexandra Meliou	
CXL Memory Performance for In-Memory Data Processing.....	3119
Marcel Weisgut, Daniel Ritter, Pinar Tözün, Lawrence Benson, Tilmann Rabl	
LLMLog: Advanced Log Template Generation via LLM-driven Multi-Round Annotation	3134
Fei Teng, Haoyang Li, Lei Chen	
Cuckoo Heavy Keeper and the balancing act of maintaining heavy hitters in stream processing	3149
Vinh Ngo, Marina Papatriantafidou	

Rebirth-Retire: A Concurrency Control Protocol Adaptable to Different Levels of Contention	3162
Qian Zhang, Yiwen Xiang, Jianhao Wei, Yang Yang, Yifan Li, Xueqing Gong, Wanggen Liu	
\textsc{Ufgtime}: Mining Intertwined Dependencies in Multivariate Time Series via an Efficient Pure Graph Approach (Flavor: Foundations and Algorithms Papers)	3175
Ruikun Li, Dai Shi, Ye Xiao, Junbin Gao	
ArrayMorph: Optimizing Hyperslab Queries on the Cloud for Machine Learning Pipelines	3189
Ruochen Jiang, Spyros Blanas	
Inference-friendly Graph Compression for Graph Neural Networks.....	3203
Yangxin Fan, Haolai Che, Yinghui Wu	
GpJSON: High-performance JSON Data Processing on GPUs.....	3216
Sacheendra Talluri, Guido Walter Di Donato, Luca Danelutti, Koen Vlaswinkel, Marco Arnaboldi, Arnaud Delamare, Marco Domenico Santambrogio, Daniele Bonetta	
Beyond Shortest Paths: Node Fairness in Route Recommendation.....	3230
Antonio Ferrara, David Garcia-Soriano, Francesco Bonchi	

TABLE OF CONTENTS

Research Papers

Access Control for Information-Theoretically Secure Data.....	3243
Yin Li, Sharad Mehrotra, Shantanu Sharma, Komal Kumari	
Dynamic Range-Filtering Approximate Nearest Neighbor Search	3256
Zhencan Peng, Miao Qiao, Wenchao Zhou, Feifei Li, Dong Deng	
LEGO-GraphRAG: Modularizing Graph-based Retrieval-Augmented Generation for Design Space Exploration.....	3269
Yukun Cao, Zengyi Gao, Zhiyang Li, Xike Xie, S. Kevin Zhou, Jianliang Xu	
Déjà Vu: Efficient Video-Language Query Engine with Learning-based Inter-Frame Computation Reuse	3284
Jinwoo Hwang, Daeun Kim, Sangyeop Lee, Yoonsung Kim, Guseul Heo, Hojoon Kim, Yunseok Jeong, Tadiwos Meaza, Eunhyeok Park, Jeongseob Ahn, Jongse Park	
Parachute: Single-Pass Bi-Directional Information Passing	3299
Mihail Stoian, Andreas Zimmerer, Skander Krid, Amadou Ngom, Jialin Ding, Tim Kraska, Andreas Kipf	
Twisted Twin: A Collaborative and Competitive Memory Management Approach in HTAP Systems	3312
Jiani Yang, Sai Wu, Yong Wang, Dongxiang Zhang, Yifei Liu, Xiu Tang, Gang Chen	
Customization Meets 2-Hop Labeling: Efficient Routing in Road Networks.....	3326
Muhammad Farhan, Henning Koehler, Qing Wang, Jiawen Wang, Moritz Laupichler, Peter Sanders	
X-Blossom: Massive Parallelization of Graph Maximum Matching.....	3339
Dayi Fan, Rubao Lee, Xiaodong Zhang	
Data Imputation with Limited Data Redundancy Using Data Lakes	3354
Chenyu Yang, Yuyu Luo, Chuanxuan Cui, Ju Fan, Chengliang Chai, Nan Tang	
Chimera: Mitigating Ownership Transfers in Multi-Primary Shared-Storage Cloud-Native Databases....	3368
Huang Chunyue, Shuang Liu, Xinyi Zhang, Wenhao Li, Wei Lu, Xiaoyong Du	
Sectric: Towards Accurate, Privacy-preserving and Efficient Triangle Counting.....	3382
Minze Xu, Zhentai Xie, Zhibin Wang, Guangzhan Wang, Longbin Lai, Yuan Zhang, Chen Tian, Sheng Zhong	
When Speed meets Accuracy: an Efficient and Effective Graph Model for Temporal Link Prediction	3396
Haoyang Li, Yuming Xu, Yiming Li, Hanmo Liu, Darian Li, Chen Jason Zhang, Lei Chen, Qing Li	
Improving Time Series Data Compression in Apache IoTDB	3406
Yuxin Tang, Feng Zhang, Jiawei Guan, Yuan Tian, Xiangdong Huang, Chen Wang, Jianmin Wang, Xiaoyong Du	
On LLM-Enhanced Mixed-Type Data Imputation with High-Order Message Passing.....	3421
Jianwei Wang, Kai Wang, Ying Zhang, Wenjie Zhang, Xiwei Xu, Xuemin Lin	

Meaningful Data Erasure in the Presence of Dependencies.....	3435
Vishal Chakraborty, Youri Kaminsky, Sharad Mehrotra, Felix Naumann, Faisal Nawab, Primal Pappachan, Mohammad Sadoghi, Nalini Venkatasubramanian	
Sonata: Multi-Database Transactions Made Fast and Serializable.....	3449
Chuzhe Tang, Zhaoguo Wang, Jinyang Li, Haibo Chen	
MOMENTI: Scalable Motif Mining in Multidimensional Time Series	3463
Matteo Ceccarello, Francesco Pio Monaco, Francesco Silvestri	
How and Why False Denial Constraints are Discovered.....	3477
Albert Martin, Eduardo C. De Almeida, Oscar Romero, Anna Queralt	
Efficient k -Clique Densest Subgraph Discovery: Towards Bridging Practice and Theory	3490
Yingli Zhou, Qingshuo Guo, Yixiang Fang	
AutoPrep: Natural Language Question-Aware Data Preparation with a Multi-Agent Framework.....	3504
Meihao Fan, Ju Fan, Nan Tang, Lei Cao, Guoliang Li, Xiaoyong Du	
Accelerating Approximate Nearest Neighbor Search in Hierarchical Graphs: Efficient Level Navigation with Shortcuts	3518
Zengyang Gong, Yuxiang Zeng, Lei Chen	
Federated Incomplete Tabular Data Prediction with Missing Complementarity.....	3531
Yan Zhang, Shuwei Liang, Xiaoye Miao, Yangyang Wu, Jianwei Yin	
LakeVisage: Towards Scalable, Flexible and Interactive Visualization Recommendation for Data Discovery over Data Lakes.....	3545
Yihao Hu, Jin Wang, Sajjadur Rahman	
PS-MI: Accurate, Efficient, and Private Data Valuation in Vertical Federated Learning.....	3559
Xiaokai Zhou, Xiao Yan, Fangcheng Fu, Ziwen Fu, Tiejun Qian, Yuanyuan Zhu, Qinbo Zhang, Bin Cui, Jiawei Jiang	
Towards Pattern-aware Data Augmentation for Temporal Knowledge Graph Completion	3573
Jiasheng Zhang, Deqiang Ouyang, Shuang Liang, Jie Shao	
RapidStore: An Efficient Dynamic Graph Storage System for Concurrent Queries	3587
Chiyu Hao, Jixian Su, Shixuan Sun, Hao Zhang, Sen Gao, Jianwen Zhao, Chenyi Zhang, Jieru Zhao, Chen Chen, Minyi Guo	
GORAM: Graph-oriented ORAM for Efficient Ego-centric Queries on Federated Graphs.....	3601
Xiaoyu Fan, Kun Chen, Jiping Yu, Xiaowei Zhu, Yunyi Chen, Huanchen Zhang, Wei Xu	
Authenticated Aggregate Queries with Boolean Range Predicates on Blockchains	3615
Weijie Sun, Zihuan Xu, Wangze Ni, Lei Chen, Peng Cheng, Chen Jason Zhang	
FSMDTW: A Fast Index-free Subsequence Matching Algorithm for Dynamic Time Warping	3628
Zemin Chao, Qiaoyi Zheng, Zhixin Qi, Hongzhi Wang	
Fused Gromov-Wasserstein Alignment for Graph Edit Distance Computation and Beyond	3641
Jianheng Tang, Xi Zhao, Lemin Kong, Xiaofang Zhou, Jia Li	
EVOSHEMA: TOWARDS TEXT-TO-SQL ROBUSTNESS AGAINST SCHEMA EVOLUTION	3655

Tianshu Zhang, Kun Qian, Siddhartha Sahai, Yuan Tian, Shaddy Garg, Huan Sun, Yunyao Li

Effective and Efficient Community Search for Complex Network Semantics Capture: From Coarse-Grain to Fine-Grain 3669
Shuai Han, Yushi Tao, Jingwen Tan, Huanran Wang, Wu Yang, Yanmei Wang

HAWK: A Workload-driven Hierarchical Deadlock Detection Approach in Distributed Database System
..... 3682
Rongrong Zhang, Zhiwei Ye, Jun-Peng Zhu, Peng Cai, Xuan Zhou, Dunbo Cai, Ling Qian

TABLE OF CONTENTS

Research Papers

Doctopus: Budget-aware Structural Table Extraction from Unstructured Documents.....	3695
Chengliang Chai, Jiajun Li, Yuhao Deng, Yuanhao Zhong, Ye Yuan, Guoren Wang, Lei Cao	
S^3AND: Efficient Subgraph Similarity Search Under Aggregated Neighbor Difference Semantics.....	3708
Qi Wen, Yutong Ye, Xiang Lian, Mingsong Chen	
Lighter-X: An Efficient and Plug-and-play Strategy for Graph-based Recommendation through Decoupled Propagation.....	3721
Yanping Zheng, Zhewei Wei, Frank De Hoo, Xu Chen, Hongteng Xu, Yuhang Ye, Jiadeng Huang	
Not Small Enough? SegPQ: A Learned Approach to Compress Product Quantization Codebooks.....	3730
Qiyu Liu, Yanlin Qi, Siyuan Han, Jingshu Peng, Jin Li, Lei Chen	
The Accuracy of Cardinality Estimators: Unraveling the Evaluation Result Conundrum	3744
Nazanin Rashedi, Guido Moerkotte	
LogLite: Lightweight Plug-and-Play Streaming Log Compression.....	3757
Benzhao Tang, Shiyu Yang, Zhitao Shen, Wenjie Zhang, Xuemin Lin, Zhihong Tian	
Enabling Efficient Attack Investigation via Human-in-the-Loop Security Analysis	3771
Saimon Tsegai, Xinyu Yang, Haoyuan Liu, Peng Gao	
Shifting Transaction Isolation on Graphs: From Systems to Data.....	3784
Wenzhi Fu, Yang Cao	
Fast Graph Vector Search via Hardware Acceleration and Delayed-Synchronization Traversal	3797
Wenqi Jiang, Hang Hu, Torsten Hoefler, Gustavo Alonso	
Fremer: Lightweight and Effective Frequency Transformer for Workload Forecasting in Cloud Services	3812
Hengyu Ye, Jiadong Chen, Xiao He, Fuxin Jiang, Tieying Zhang, Jianjun Chen, Xiaofeng Gao	
TabulaX: Leveraging Large Language Models for Multi-Class Table Transformations	3826
Arash Dargahi Nobari, Davood Rafiei	
Bonspiel: Low Tail Latency Transactions in Geo-Distributed Databases	3840
Fan Cui, Eric Lo, Srijan Srivastava, Ziliang Lai	
Efficient Graph Data Access for Out-of-Memory GPU Streaming Graph Processing	3854
Qiange Wang, Yongze Yan, Hongshi Tan, Cheng Chen, Cheng Zhao, Jiaming Tian, Jiaxin Jiang, Xiaoliang Cong, Yanfeng Zhang, Ge Yu, Weng-Fai Wong, Bingsheng He	
Extensible and Robust Evaluation of Similarity Queries	3868
Daniel Schmitt, Thomas Hütter, Nikolaus Augsten	

PBench: Workload Synthesizer with Real Statistics for Cloud Analytics Benchmarking	3883
Yan Zhou, Chunwei Liu, Bhuvan Uргаonkar, Zhengle Wang, Magnus Mueller, Chao Zhang, Songyue Zhang, Pascal Pfeil, Dominik Horn, Zhengchun Liu, Davide Pagano, Tim Kraska, Samuel Madden, Ju Fan	
Accelerating Subgraph Matching through Fine-grained and Powerful Equivalences	3896
Yujie Lu, Zhijie Zhang, Weiguo Zheng, Lei Zou	
How to Optimize SQL Queries? A Comparison Between Split, Holistic, and Hybrid Approaches	3910
Luca Gretscher, Jens Dittrich	
Diva: Dynamic Range Filter for Var-Length Keys and Queries	3923
Navid Eslami, Ioana Bercea, Niv Dayan	
Approximate 2-hop neighborhoods on incremental graphs: An efficient lazy approach	3937
Luca Becchetti, Andrea Clementi, Luciano Gualà, Luca Pepè Sciarria, Alessandro Straziota, Matteo Stromieri	
Cracking Vector Search Indexes.....	3951
Vasilis Mageirakos, Bowen Wu, Gustavo Alonso	
DobLIX: A Dual-Objective Learned Index for Log-Structured Merge Trees	3965
Alireza Heidari, Amirhossein Ahmadi, Wei Zhang	
CoLA: Model Collaboration for Log-based Anomaly Detection	3979
Xuhang Zhu, Xiu Tang, Sai Wu, Jichen Li, Haobo Wang, Chang Yao, Quanqing Xu, Gang Chen	
Towards Designing Future-Proof Data Processing Systems.....	3988
Michael Jungmair, Jana Giceva	
Advancing Fact Attribution for Query Answering: Aggregate Queries and Novel Algorithms	3996
Omer Abramovich, Daniel Deutch, Nave Frost, Ahmet Kara, Dan Olteanu	
What If: Causal Analysis with Graph Databases	4009
Amedeo Pachera, Mattia Palmiotto, Angela Bonifati, Andrea Mauri	
AnyBlox: A Framework for Self-Decoding Datasets.....	4017
Mateusz Gieniecъzko, Maximilian Kuschewski, Thomas Neumann, Viktor Leis, Jana Giceva	
Federated and Balanced Clustering for High-dimensional Data.....	4032
Yushuai Ji, Shengkun Zhu, Shixun Huang, Zepeng Liu, Sheng Wang, Zhiyong Peng	
Relational Data Models for Genetic VCF data	4045
Mohamed Sabri Hafidi, Ozan Kahramanođulları, Anton Dignös, Johann Gamper	
BURST: Rendering Clustering Techniques Suitable for Evolving Streams.....	4054
Apostolos Giannoulidis, Anastasios Gounaris, John Paparrizos	
Environmental Footprints of Query Processing: A Vision for Sustainable Database Architectures.....	4064
Michail Bachras, Hans-Arno Jacobsen	
Semantic Integrity Constraints: Declarative Guardrails for AI-Augmented Data Processing Systems.....	4073
Alexander Lee, Justin Chan, Michael Fu, Nicolas Kim, Akshay Mehta, Deepti Raghavan, Ugur Cetintemel	
RICH: Real-time Identification of negative Cycles for High-efficiency Arbitrage.....	4081

Bingqiao Luo, Jiaxin Jiang, Yuhang Chen, Junyi Hou, Cheng Jun Tey, Ziyang Qiu, Bingsheng He, Spencer Xiao, Dominic Ong, Wee Howe Ang

Balancing Privacy and Utility in Correlated Data: A Study of Bayesian Differential Privacy 4090
Martin Lange, Patricia Guerra-Balboa, Javier Parra-Arnau, Thorsten Strufe

Enhancing Transaction Processing through Indirection Skipping 4104
Riki Otaki, Jun Hyuk Chang, Aaron Elmore, Goetz Graefe

UniClean: A Scalable Data Cleaning Solution for Mixed Errors based on Unified Cleaners and Optimized Cleaning Workflow 4117
Xiaoou Ding, Zekai Qian, Hongzhi Wang, Siying Chen, Yafeng Tang, Hongbin Su, Huan Hu, Chen Wang

ShaRP: Explaining Rankings and Preferences with Shapley Values 4131
Venetia Pliatsika, Joao Fonseca, Kateryna Akhynko, Ivan Shevchenko, Julia Stoyanovich

SQLStorm: Taking Database Benchmarking into the LLM Era 4144
Tobias Schmidt, Leis Viktor, Peter Boncz, Thomas Neumann

Suna: Scalable Causal Confounder Discovery over Relational Data 4158
Jiaxiang Liu, Siyuan Xia, Daniel Alabi, Eugene Wu

Semantic Operators and Their Optimization: Towards AI-Based Data Analytics with Accuracy Guarantees 4171
Liana Patel, Siddharth Jha, Melissa Pan, Harshit Gupta, Parth Asawa, Carlos Guestrin, Matei Zaharia

Improving DBMS Scheduling Decisions with Accurate Performance Prediction on Concurrent Queries 4185
Ziniu Wu, Markos Markakis, Chunwei Liu, Peter Chen, Balakrishnan Narayanaswamy, Tim Kraska, Samuel Madden

Practical and Accurate Local Edge Differentially Private Graph Algorithms 4199
Pranay Mundra, Charalampos Papamanthou, Julian Shun, Quanquan Liu

Continuous Publication of Weighted Graphs with Local Differential Privacy 4214
Wen Xu, Pengpeng Qiao, Shang Liu, Zhirun Zheng, Yang Cao, Zhetao Li

TxnSails: Achieving Serializable Transaction Scheduling with Self-Adaptive Isolation Level Selection... 4227
Qiyu Zhuang, Wei Lu, Shuang Liu, Yuxing Chen, Xinyue Shi, Zhanhao Zhao, Yipeng Sun, Anqun Pan, Xiaoyong Du

No Cap, This Memory Slaps: Breaking Through the Memory Wall of Transactional Database Systems with Processing-in-Memory 4241
Hyounghoo Kim, Yiwei Zhao, Andrew Pavlo, Phillip B. Gibbons

GraphCSR: A Degree-Equalized CSR Format for Large-scale Graph Processing 4255
Xinbiao Gan, Tiejun Li, Chunye Gong, Dongsheng Li, Dezun Dong, Jie Liu, Kai Lu

Effective and Efficient Attributed Hypergraph Embedding on Nodes and Hyperedges 4269
Yiran Li, Gongyao Guo, Chen Feng, Jieming Shi

Subgraph Matching: A New Decomposition Based Approach 4282
Qiyang Li, Jeffrey Yu, Zongyan He

SSD-iq: Uncovering the Hidden Side of SSD Performance 4295
Gabriel Haas, Bohyun Lee, Philippe Bonnet, Viktor Leis

Faster Convergence in Mini-batch Graph Neural Networks Training with Pseudo Full Neighborhood Compensation.....	4309
Qiqi Zhou, Yanyan Shen, Lei Chen	
TreeCat: Standalone Catalog Engine for Large Data Systems.....	4323
Keonwoo Oh, Pooja Nilangekar, Amol Deshpande	
Select Edges Wisely: Monotonic Path Aware Graph Layout Optimization for Disk-based ANN Search....	4337
Ziyang Yue, Bolong Zheng, Ling Xu, Kanru Xu, Shuhao Zhang, Yajuan Du, Yunjun Gao, Xiaofang Zhou, Christian Jensen	
Powerful GPUs or Fast Interconnects: Analyzing Relational Workloads on Modern GPUs.....	4350
Marko Kabić, Bowen Wu, Jonas Dann, Gustavo Alonso	
TSB-AutoAD: Towards Automated Solutions for Time-Series Anomaly Detection.....	4364
Qinghua Liu, Seunghak Lee, Paparrizos John	
Time-Series Clustering: A Comprehensive Study of Data Mining, Machine Learning, and Deep Learning Methods.....	4380
John Paparrizos, Sai Prasanna Teja Reddy Bogireddy	
Beyond Compression: A Comprehensive Evaluation of Lossless Floating-Point Compression.....	4396
Kaisei Hishida, Chunwei Liu, John Paparrizos, Aaron Elmore	
ThriftLLM: On Cost-Effective Selection of Large Language Models for Classification Queries.....	4410
Keke Huang, Yimin Shi, Dujian Ding, Yifei Li, Yang Fei, Laks Lakshmanan, Xiaokui Xiao	
Sphinx: A Succinct Perfect Hash Index for x86.....	4424
Sajad Faghfoor Maghrebi, Niv Dayan	
NaviX: A Native Vector Index Design for Graph DBMSs With Robust Predicate-Agnostic Search Performance.....	4438
Gaurav Sehgal, Semih Salihoglu	
DIM-SUM: Dynamic IMputation for Smart Utility Management.....	4451
Ryan Hildebrant, Rahul Bhope, Sharad Mehrotra, Christopher Tull, Nalini Venkatasubramanian	
Robust Recursive Query Parallelism in Graph Database Management Systems.....	4465
Anurag Chakraborty, Semih Salihoglu	
OasisDB: An Oblivious and Scalable System for Relational Data.....	4478
Haseeb Ahmed, Nachiket Rao, Abdelkarim Kati, Florian Kerschbaum, Sujayya Maiyya	
CEDAR: A System for Cost-Efficient Data-Driven Claim Verification.....	4492
Tharushi Jayasekara, Immanuel Trummer	
Benchmarking Adaptive Multidimensional Indices.....	4505
Konstantinos Lampropoulos, Fatemeh Zardbani, Nikos Mamoulis, Panagiotis Karras	
Scaling GPU-Accelerated Databases beyond GPU Memory Size.....	4518
Yinan Li, Bailu Ding, Ziyun Wei, Lukas Maas, Momin Al-Ghosien, Spyros Blanas, Nicolas Bruno, Carlo Curino, Matteo Interlandi, Craig Peepers, Kaushik Rajan, Surajit Chaudhuri, Johannes Gehrke	
PAR2QO: Parametric Penalty-Aware Robust Query Optimization.....	4532

Haibo Xiu, Yang Li, Qianyu Yang, Pankaj Agarwal, Jun Yang

LIMAO: A Framework for Lifelong Modular Learned Query Optimization..... 4546
Qihan Zhang, Shaolin Xie, Ibrahim Sabek

QUEST: Query Optimization in Unstructured Document Analysis 4560
Zhaoze Sun, Chengliang Chai, Qiyang Deng, Kaisen Jin, Xinyu Guo, Han Han, Ye Yuan, Guoren Wang, Lei Cao

CENTS: A Flexible and Cost-Effective Framework for LLM-Based Table Understanding..... 4574
Guorui Xiao, Dong He, Jin Wang, Magdalena Balazinska

OmniMatch: Joinability Discovery in Data Products 4588
Christos Koutras, Jiani Zhang, Xiao Qin, Chuan Lei, Vassilis Ioannidis, Christos Faloutsos, George Karypis, Asterios Katsifodimos

Pistis: A Decentralized Knowledge Graph Platform Enabling Ownership-Preserving SPARQL Querying..... 4602
Enyuan Zhou, Song Guo, Zicong Hong, Christian Jensen, Yang Xiao, Jinwen Liang, Dalin Zhang

Selective Late Materialization in Modern Analytical Databases 4616
Yihao Liu, Shaoxuan Tang, Yulong Hui, Hangrui Zhou, Huanchen Zhang

The FastLanes File Format 4629
Azim Afroozeh, Peter Boncz

POLARIS: An Interactive and Scalable Data Infrastructure for Polar Science 4644
Yuchuan Huang, Ana Elena Uribe, Kareem Eldahshoury, Youssef Hussein, Grant Ogren, Mohamed Mokbel

Efficiently Joining Large Relations on Multi-GPU Systems 4653
Tobias Maltnerberger, Ilin Tolovski, Tilmann Rabl

Stress-Testing ML Pipelines with Adversarial Data Corruption 4668
Jiongli Zhu, Geyang Xu, Felipe Lorenzi, Boris Glavic, Babak Salimi

PrivAGM: Secure Construction of Differentially Private Directed Attributed Graph Models on Decentralized Social Graphs..... 4682
Songlei Wang, Yifeng Zheng, Xiaohua Jia, Haibo Hu

OmniSQL: Synthesizing High-quality Text-to-SQL Data at Scale..... 4695
Haoyang Li, Shang Wu, Xiaokang Zhang, Xinmei Huang, Jing Zhang, Fuxin Jiang, Shuai Wang, Tieying Zhang, Jianjun Chen, Rui Shi, Hong Chen, Cuiping Li

Turbocharging Vector Databases using Modern SSDs..... 4710
Joobo Shim, Jaewon Oh, Hongchan Roh, Jaeyoung Do, Sang-Won Lee

SIEVE: Effective Filtered Vector Search with Collection of Indexes 4723
Zhaoheng Li, Silu Huang, Wei Ding, Yongjoo Park, Jianjun Chen

Enhancing Graph Edit Distance Computation: Stronger and Orientation-based ILP Formulations..... 4737
Andrea D'ascenzo, Julian Meffert, Petra Mutzel, Fabrizio Rossi

ParSEval: Plan-aware Test Database Generation for SQL Equivalence Evaluation 4750
Chunyu Chen, Zhengjie Miao, Yong Zhang, Jiannan Wang

TABLE OF CONTENTS

Industrial Papers

LEADRE: Multi-Faceted Knowledge Enhanced LLM Empowered Display Advertisement Recommender System	4763
<i>Fengxin Li, Yi Li, Yue Liu, Chao Zhou, Yuan Wang, Xiaoxiang Deng, Wei Xue, Dapeng Liu, Lei Xiao, Haijie Gu, Jie Jiang, Hongyan Liu, Biao Qin, Jun He</i>	
TuskFlow: An Efficient Graph Database for Long-Running Transactions.....	4777
<i>Georgios Theodorakis, Hugo Firth, James Clarkson, Natacha Crooks, Jim Webber</i>	
MD-MVCC: Multi-version Concurrency Control for Schema Changes in Azure SQL Database	4791
<i>Panagiotis Antonopoulos, Mansi Chauhan, Shailender Dabas, Rajat Jain, Darshan Kattera, Wonseok Kim, Hanuma Kodavalla, Nikolas Ogg, Prashanth Purnananda, Rahul Ranjan, Alex Swanson, Divyesh Tikmani</i>	
Towards Principled, Practical Document Database Design	4804
<i>Michael Carey, Wail Alkowiileet, Nick Digeronimo, Peeyush Gupta, Sachin Smotra, Till Westmann</i>	
Scribe: How Meta transports zettabytes per day in real time	4817
<i>Manos Karpathiotakis, Vlassios Rizopoulos, Artem Gelun, Tiziano Carotti, Hazem Nada, Basri Kahveci, Yuri Dolgov</i>	
The HANA Native Query Engine for Lakehouse Systems	4831
<i>Daniel Ritter, Mihnea Andrei, Sukhyeun Cho, Maik Goergens, Taehyung Lee, Norman May, Amit Pathak, Paul Willems</i>	
Disaggregated State Management in Apache Flink 2.0	4846
<i>Yuan Mei, Zhaoqian Lan, Lei Huang, Yanfei Lei, Han Yin, Rui Xia, Kaitian Hu, Paris Carbone, Vasiliki Kalavri, Feng Wang</i>	
SiriusBI: A Comprehensive LLM-powered Solution for Data Analytics in Business Intelligence	4860
<i>Jie Jiang, Haining Xie, Siqi Shen, Yu Shen, Zihan Zhang, Meng Lei, Yifeng Zheng, Yang Li, Chunyou Li, Danqing Huang, Yinjun Wu, Wentao Zhang, Xiaofeng Yang, Bin Cui, Peng Chen</i>	
SagaLLM: Context Management, Validation, and Transaction Guarantees for Multi-Agent LLM Planning	4874
<i>Edward Chang, Longling Geng</i>	
Unlocking the Power of CI/CD for Data Pipelines in Distributed Data Warehouses.....	4887
<i>Hongtao Yang, Zhichen Xu, Sergey Yudin, Andrew Davidson</i>	
veDB-HTAP: a Highly Integrated, Efficient and Adaptive HTAP System.....	4896

Jianjun Chen, Li Zhang, Yu Xie, Wei Ding, Lixun Cao, Ye Liu, Yonghua Ding, Fangshi Li, Ke Wu, Haibo Xiu, Kui Wei, Le Cai, Rui Chang, Yuxiang Chen, Yuanjin Lin, Shangyu Luo, Jianfeng Qian, Xu Wang, Zikang Wang, Jian Zhang, Mingyi Zhang, Shicai Zeng, Jason Sun, Lei Zhang, Rui Shi, Pengwei Zhao

From FASTER to F2: Evolving Concurrent Key-Value Store Designs for Large Skewed Workloads..... 4910
Konstantinos Kanellis, Badrish Chandramouli, Ted Hart, Shivaram Venkataraman

Automatic Indexing in Oracle..... 4924
Sunil Chakkappen, Shreya Kunjibettu, Daniel Mcgreer, Masoomah Kishi, Hong Su, Mohamed Ziauddin, Mohamed Zait, Zhan Li, Yuying Zhang

Cloudy With a Chance of JSON 4938
Murtadha Al Hubail, Ali Alsuliman, Wail Alkwaileet, Michael Blow, Michael Carey, Savyasach Erukonda, Peeyush Gupta, Santosh Hegde, Kamini Jagtiani, Abhishek Jindal, Nawazish Kahn, Mehnaz Mahin, Ian Maxon, M Muralikrishna, Keshav Murthy, Daniel Nagy, Preetham Poluparthi, Ankit Prabhu, Ritik Raj, Vijay Sarathy, Shahrzad Shirazi, Utsav Singh, Hussain Towaileb, Ayush Tripathi, Janhavi Tripurwar, Bo-Chun Wang, Till Westmann

GaussDB-Vector: A Large-Scale Persistent Real-Time Vector Database for LLM Applications 4951
Guoliang Li, Ji Sun, James Pan, Jiang Wang, Yongqing Xie, Ruicheng Liu, Wen Nie

Magnus: A Holistic Approach to Data Management for Large-Scale Machine Learning Workloads 4964
Jun Song, Jingyi Ding, Irshad Kandy, Yanghao Lin, Zhongjia Wei, Zilong Zhou, Zhiwei Peng, Jixi Shan, Hongyue Mao, Xiuqi Huang, Xun Song, Cheng Chen, Yanjia Li, Tianhao Yang, Wei Jia, Xiaohong Dong, Kang Lei, Rui Shi, Pengwei Zhao, Wei Chen

DECK: Experiences on Delta Checkpointing for Industrial Recommendation Systems 4978
Xin Gao, Sibasish Acharya, Sihui Han, Yongxiong Ren, Yanli Zhao, Liang Luo, Chucheng Wang, Pradeep Fernando, Saurabh Mishra, Siqi Yan, Yicong Du, Elzbieta Krepska, Intaik Park, Min Ni, Qunshu Zhang, Shen Li

GRewriter: Practical Query Rewriting with Automatic Rule Set Expansion in GaussDB 4991
Zhe Jiang, Zhaoguo Wang, Haoning Lan, Chuzhe Tang, Haoran Ding, Lefeng Wang, Songyun Zou, Zhuoran Wei, Yongcun Liu, Xiang Yu, Yang Ren, Guoliang Li, Haibo Chen

FDBKeeper: Enabling Scalable Coordination Services for Metadata Management using Distributed Key-Value Databases..... 5004
Jun-Peng Zhu, Lingfeng Zhang, Peng Cai, Xuan Zhou, Peisen Zhao, Xue Wang, Linpeng Tang

VSAG: An Optimized Search Framework for Graph-based Approximate Nearest Neighbor Search..... 5017
Xiaoyao Zhong, Haotian Li, Jiabao Jin, Mingyu Yang, Deming Chu, Xiangyu Wang, Zhitao Shen, Wei Jia, George Gu, Yi Xie, Xuemin Lin, Heng Tao Shen, Jingkuan Song, Peng Cheng

R-Bot: An LLM-based Query Rewrite System..... 5031
Zhaoyan Sun, Xuanhe Zhou, Guoliang Li, Xiang Yu, Jianhua Feng, Yong Zhang

Design and Modular Verification of Distributed Transactions in MongoDB 5045
William Schultz, Murat Demirbas

From Scale-Up to Scale-Out: PolarDB's Journey to Achieving 2 Billion tpmC 5059
Xinjun Yang, Feifei Li, Yingqiang Zhang, Hao Chen, Qingda Hu, Panfeng Zhou, Qiang Zhang, Shuai Li, Zongzhi Chen, Zheyu Miao, Rongbiao Xie, Chuan Sun, Zetao Wei, Jing Fang, Xingxuan Zhou, Xiaofei Wu

ScaleCache: Scalable and Production-grade Buffer Management for Disk-based Database Systems..... 5073
Mingyu Liu, Junbin Kang, Kai Wang, Lu Zhang, Haibo Chen, Xiuchang Li, Tianhong Ding

Towards Automated Cross-domain Exploratory Data Analysis through Large Language Models	5086
<i>Jun-Peng Zhu, Boyan Niu, Peng Cai, Zheming Ni, Jianwei Wan, Kai Xu, Jiajun Huang, Shengbo Ma, Bing Wang, Xuan Zhou, Guanglei Bao, Donghui Zhang, Liu Tang, Qi Liu</i>	
GalaxyWeaver: Autonomous Table-to-Graph Conversion and Schema Optimization with Large Language Models	5100
<i>Bing Tong, Yan Zhou, Chen Zhang, Jianheng Tang, Jia Li, Lei Chen</i>	
Freely Moving Between the OLTP and OLAP Worlds: Hermes – an High-Performance OLAP Accelerator for MySQL.....	5113
<i>Tim Gubner, Rune Humborstad, Manyi Lu</i>	
Workload Insights From the Snowflake Data Cloud: What Do Production Analytic Queries Really Look Like?.....	5126
<i>Jan Vincent Szlang, Sebastian Breß, Sebastian Cattes, Jonathan Dees, Florian Funke, Max Heimel, Michel Oleynik, Ismail Oukid, Tobias Maltenberger</i>	
AnalyticDB-PG: A Cloud-native High-performance Data Warehouse in Alibaba Cloud.....	5139
<i>Fangyuan Zhang, Caihua Yin, Hua Fan, Fenghua Fang, Yineng Chen, Xuqi Wang, Mengqi Wu, Bing Chen, Tianbo Jin, Sibao Wang, Wenchao Zhou, Feifei Li</i>	
Streaming View: An Efficient Data Processing Engine for Modern Real-time Data Warehouse of Alibaba Cloud	5153
<i>Fangyuan Zhang, Mengqi Wu, Chunlei Xu, Yunong Bao, Jiyu Qiao, Yingli Zhou, Hua Fan, Caihua Yin, Wenchao Zhou, Feifei Li</i>	
Cost-Effective, Low Latency Vector Search with Azure Cosmos DB.....	5166
<i>Nitish Upreti, Harsha Simhadri, Hari Sundar, Krishnan Sundaram, Samer Boshra, Bala Perumalswamy, Shivam Atri, Martin Chisholm, Revti Singh, Greg Yang, Tamara Hass, Nitesh Dudhey, Subramanyam Pattipaka, Mark Hildebrand, Magdalen Manohar, Jack Moffitt, Haiyang Xu, Naren Datha, Suryansh Gupta, Ravi Krishnaswamy, Prashant Gupta, Abhishek Sahu, Hemeswari Varada, Sudhanshu Barthwal, Ritika Mor, James Codella, Shaun Cooper, Kevin Pilch, Simon Moreno, Aayush Kataria, Santosh Kulkarni, Neil Deshpande, Amar Sagare, Dinesh Billa, Zishan Fu, Vipul Vishal</i>	
Ursa: A Lakehouse-Native Data Streaming Engine for Kafka	5184
<i>Sijie Guo, Matteo Merli, Hang Chen, Neng Lu, Penghui Li</i>	
Delta Sharing: An Open Protocol for Cross-Platform Data Sharing.....	5197
<i>Krishna Puttaswamy, Abhijit Chakankar, Tao Tao, Zaheera Valani, Ramesh Chandra, William Chau, Mengxi Chen, Akram Chetibi, Tianyi Huang, Jonathan Keller, Celia Kung, Andy Liu, Charlene Lyu, Samarth Shetty, Xiaotong Sun, Steve Weis, Lin Zhou, Ryan Zhu, Reynold Xin, Matei Zaharia</i>	
SQL:Trek Automated Index Design at Airbnb	5210
<i>Sam Lightstone, Ping Wang</i>	

Demonstrations

TCO2: Analyzing the Carbon Footprint of Database Server Replacements	5223
<i>Marc Baeuerle, Thomas Bodner, Martin Boissier, Tilmann Rabl, Ricardo Salazar Diaz, Florian Schmeller, Nils Straßenburg, Ilin Tolovski, Marcel Weisgut, Wang Yue</i>	
FDepHunter: Harnessing Negative Examples to Expose Fakes and Reveal Ghosts.....	5227
<i>Pavel Koupil, Jáchym Bártík, Stefan Klessinger, André Conrad, Stefanie Scherzinger</i>	

VIDEX: A Disaggregated and Extensible Virtual Index for Cloud-Native and AI-Driven Databases.....	5231
<i>Rong Kang, Shuai Wang, Tieying Zhang, Xianghong Xu, Linhui Xu, Zhimin Liang, Lei Zhang, Rui Shi, Jianjun Chen</i>	
DVote: Constraining Committee Voting with Database Dependencies.....	5235
<i>Roi Yona, Jonathan Breitman, Benny Kimelfeld</i>	
Graph Compression for Interpretable Graph Neural Network Inference At Scale.....	5239
<i>Yangxin Fan, Haolai Che, Mingjian Lu, Yinghui Wu</i>	
Query running too slow? Rewrite it with Quorion!	5243
<i>Bingnan Chen, Binyang Dai, Qichen Wang, Ke Yi</i>	
Demonstration of ModelarDB: Model-Based Management of High-Frequency Time Series Across Edge, Cloud, and Client.....	5247
<i>Søren Kejser Jensen, Christian Schmidt Godiksen, Christian Thomsen, Torben Bach Pedersen</i>	
Democratize MATCH_RECOGNIZE!	5251
<i>Louisa Lambrecht, Tim Findling, Samuel Heid, Marcel Knüdel, Torsten Grust</i>	
Opening The Black-Box: Explaining Learned Cost Models For Databases.....	5255
<i>Roman Heinrich, Oleksandr Havrylov, Manisha Luthra, Johannes Wehrstein, Carsten Binnig</i>	
Horizon: Robust Checks for SQL Migration Using LLMs	5259
<i>Venkatesh Emani, Wenjing Wang, Zi Ye, Jia He, Neel Ball, Kumaraswamy Boora, Carlo Curino, Avriela Floratou, Manan Goenka, Paridhi Gupta, Vivek Gupta, Katherine Lin, Nick Litombe, Jared Meade, Suryakant Mutnal, Raghu Ramakrishnan, Sudhir Raparla, Dhruv Relwani, Shyam Sai, Vaibhave Sekar, Roneet Shaw, Harmeet Singh, Prasanna Sridharan, Mark Taylor, Sunidhi Tiwari, Yiwen Zhu</i>	
A Demonstration of QueryArtisan: Real-Time Data Lake Analysis via Dynamically Generated Data Manipulation Code	5263
<i>Wenhao Liu, Xiu Tang, Sai Wu, Chang Yao, Gongsheng Yuan, Gang Chen</i>	
RecForUS: A Recommender System for Uncertain Scores.....	5267
<i>Dvir Cohen, Liad Domb, Avigdor Gal, Lior Ganon, Eliezer Gavriel, Omri Lazover, Coral Scharf, Bar Shterenberg</i>	
PrivEval: a tool for interactive evaluation of privacy metrics in synthetic data generation	5271
<i>Frederik Trudslev, Matteo Lissandrini, Juan Rodriguez, Martin Bøgsted, Daniele Dell'aglio</i>	
Styx in Action: Transactional Cloud Applications Made Easy	5275
<i>Kyriakos Psarakis, Oto Mraz, George Christodoulou, George Siachamis, Marios Fragkoulis, Asterios Katsifodimos</i>	
Can Surrogate Keys Negatively Impact Data Quality?	5279
<i>Mathilde Marcy, Jean-Marc Petit, Marian Scuturici, Jocelyn Bonjour, Camille Fertel, Gerald Cavalier</i>	
JUSTINE (JUST-INsert Engine): Demonstrating Self-organizing Data Schemas.....	5283
<i>Benjamin Hättasch, Leon Krüger, Carsten Binnig</i>	
Unify: A System For Unstructured Data Analytics	5287
<i>Jiayi Wang, Yuan Li, Jianming Wu, Shihui Xu, Guoliang Li</i>	
UmbraPerf - Profiling Results Tailored for DBMS Developers	5291
<i>Alexander Beischl, Thomas Neumann</i>	

Smart SPARQL Advisor: Guiding Users in Query Formulation with Performance Prediction	5295
<i>Abiram Mohanaraj, Matteo Lissandrini, Katja Hose</i>	
Analytics Are Heavy. The DBMS Is Busy. When Will My Mission-Critical Transaction Start Running?	5299
<i>Jiatang Zhou, Kaisong Huang, Zhuoyue Zhao, Dong Xie, Tianzheng Wang</i>	
Accelerating Tabular Inference: Training Data Generation with TENET	5303
<i>Enzo Veltri, Donatello Santoro, Jean-Flavien Bussotti, Paolo Papotti</i>	
Accordion: Balancing Performance and Cost in Cloud-Native Data Analysis with Intra-Query Runtime Elasticity	5307
<i>Xukang Zhang, Huanchen Zhang, Xiaofeng Meng</i>	
Demonstration of Reflex: How SMPC Query Execution can be sped up through Efficient and Flexible Intermediate Result Size Trimming	5311
<i>Long Gu, Shaza Zeitouni, Carsten Binnig, Zsolt István</i>	
Enter the Warp: Fast and Adaptive Data Transfer with XDBC	5315
<i>Haralampos Gavriilidis, Joel Ziegler, Midhun Kaippillil Venugopalan, Benedikt Didrich, Matthias Boehm, Volker Markl</i>	
RadLER: Deduplicated Sampling On-Demand	5319
<i>Luca Zecchini, Ziawasch Abedjan, Vasilis Efthymiou, Giovanni Simonini</i>	
MLN-geeWhiz: A Dashboard for Supporting Complete Life-Cycle of Complex Data Analysis using Multilayer Networks	5323
<i>Amev Shinde, Viraj Sabhaya, Kevin Farokhrouz, Fariba Irany, Ali Khan, Sanjukta Bhowmick, Abhishek Santra, Sharma Chakravarthy</i>	
Hint-QPT: Hints for Robust Query Performance Tuning	5327
<i>Haibo Xiu, Yang Li, Qianyu Yang, Weihang Guo, Yuxi Liu, Sudeepa Roy, Pankaj Agarwal, Jun Yang</i>	
ClaimIt: Finding Convincing Views to Endorse a Claim	5331
<i>Shunit Agmon, David Avigdor, Brit Youngmann, Amir Gilad, Benny Kimelfeld</i>	
DortDB: Bridging Query Languages for Multi-Model Data Ponds	5335
<i>Filip Ježek, Pavel Koupil, Michal Kopecký, Jáchym Bártek, Irena Holubová</i>	
DemandClean: A Multi-Objective Learning Framework for Balancing Model Tolerance to Data Authenticity and Diversity	5339
<i>Zekai Qian, Xiaoou Ding, Chen Wang, Hongzhi Wang</i>	
TARImpute: Task-Aware Auto-Recommender System for Missing Value Imputation Algorithms with Clustering Case Studies	5343
<i>Xiaoou Ding, Yanshuo Liu, Zhounan Chen, Hongzhi Wang, Chen Wang, Jianmin Wang</i>	
A Demonstration of POLARIS: An Interactive and Scalable Data Infrastructure for Polar Science	5347
<i>Yuchuan Huang, Ana Elena Uribe, Grant Ogren, Youssef Hussein, Kareem Eldahshoury, Mohamed Mokbel</i>	
GooseDB: A Database Engine that Optimally Refines Top-k Queries to Satisfy Representation Constraints	5351
<i>Zixuan Chen, Jinyang Li, H. V. Jagadish, Mirek Riedewald</i>	
NeuroFlinkCEP: Neurosymbolic Complex Event Recognition Optimized across IoT Platforms	5355
<i>Ourania Ntouni, Dimitrios Banelas, Nikos Giatrakos</i>	

mlidea: Interactively Improving ML Data Preparation Code via "Shadow Pipelines"	5359
<i>Stefan Grafberger, Paul Groth, Sebastian Schelter</i>	
Mining Meaningful Keys and Foreign Keys with High Precision and Recall	5363
<i>Henning Koehler, Sebastian Link</i>	
SDG-KG: A Framework to Compute SDG Indicator using Open Data	5367
<i>Wissal Benjira, Nicolas Travers, Bénédicte Bucher, Malika Grim-Yefsah, Faten Atigui</i>	
FedVSE: A Privacy-Preserving and Efficient Vector Search Engine for Federated Databases	5371
<i>Zheng Fan, Yuxiang Zeng, Zhuanglin Zheng, Yongxin Tong</i>	
APEX-DAG: Library and Language independent Pipeline EXtraction	5375
<i>Sebastian Eggers, Nina Żukowska, Ziawasch Abedjan</i>	
Demonstrating Matelda for Multi-Table Error Detection.....	5379
<i>Fatemeh Ahmadi, Julian Paulußen, Ziawasch Abedjan</i>	
DBPecker: A Graph-Based Compound Anomaly Diagnosis System for Distributed RDBMSs.....	5383
<i>Qingliu Wu, Qingfeng Xiang, Yingxia Shao, Qiyao Luo, Quanqing Xu</i>	
DocDB: A Database for Unstructured Document Analysis	5387
<i>Zequn Li, Yuanhao Zhong, Chengliang Chai, Zhaoze Sun, Yuhao Deng, Ye Yuan, Guoren Wang, Lei Cao</i>	
ContextCache: Context-Aware Semantic Cache for Multi-Turn Queries in Large Language Models.....	5391
<i>Jianxin Yan, Wangze Ni, Lei Chen, Xuemin Lin, Peng Cheng, Zhan Qin, Kui Ren</i>	
Simulating a Transactional Server for Multi-Model Systems	5395
<i>Zenon Zacouris, Maribel Acosta</i>	
TableCopilot: A Table Assistant Empowered by Natural Language Conditional Table Discovery.....	5399
<i>Lingxi Cui, Guanyu Jiang, Huan Li, Ke Chen, Lidan Shou, Gang Chen</i>	
LETIndex: A Secure Learned Index with TEE.....	5403
<i>Shuting Cao, Zeping Niu, Guoliang Li</i>	
Play2Win: A Windowing Playground for Continuous Queries	5407
<i>Alessandro Ferri, Mauro Famà, Samuele Langhi, Riccardo Tommasini, Angela Bonifati</i>	
Vadacode: A Logician-friendly IDE for Datalog+/-	5411
<i>Luigi Bellomarini, Andrea Gentili, Davide Magnanimiti, Emanuel Sallinger</i>	
Beyond Quacking: Deep Integration of Language Models and RAG into DuckDB	5415
<i>Anas Dorbani, Sunny Yasser, Jimmy Lin, Amine Mhedhbi</i>	
SAIL: A Voyage to Symbolic Approximation Solutions for Time-Series Analysis	5419
<i>Fan Yang, John Paparrizos</i>	
Buckaroo: A Direct Manipulation Visual Data Wrangler	5423
<i>Annabelle Warner, Andrew Mcnutt, Paul Rosen, El Kindi Rezig</i>	
Sort it Like You Mean It: Discovering Semantically Interesting Attribute Augmentations to Sort Tables	5427
<i>Akash Khatri, Mahathir Mohammad, El Kindi Rezig</i>	

EasyAD: A Demonstration of Automated Solutions for Time-Series Anomaly Detection.....	5431
<i>Qinghua Liu, Seunghak Lee, John Paparrizos</i>	
LASEK: LLM-Assisted Style Exploration Kit for Geospatial Data	5435
<i>Tarlan Bahadori, Ahmed Eldawy, Sai Sreekar Sarvepalli</i>	
A Demonstration of Q2O: Quantum-augmented Query Optimizer	5439
<i>Hanwen Liu, Federico Spedalieri, Ibrahim Sabek</i>	

Tutorials

When Entity/Relationship Models Meet Graph Databases	5444
<i>Philipp Skavantzios, Sebastian Link</i>	
Synthetic Tabular Data: methods, attacks and defenses.....	5448
<i>Graham Cormode, Shripad Gade, Samuel Maddock, Enayat Ullah</i>	
Large Language Models for Spatial Analysis Queries	5451
<i>Youssef Hussein, Mohamed Hemdan, Mohamed Mokbel</i>	
Data Discovery in Data Lakes: Operations, Indexes, Systems.....	5455
<i>Ziawasch Abedjan, Mahdi Esmailoghli, Sainyam Galhorta</i>	
Systems for Scalable Graph Analytics and Machine Learning: Trends and Methods.....	5460
<i>Da Yan, Lyuheng Yuan, Akhlaque Ahmad, Saugat Adhikari</i>	
Natural Language to SQL: State of the Art and Open Problems.....	5466
<i>Yuyu Luo, Guoliang Li, Ju Fan, Chengliang Chai, Nan Tang</i>	
New Trends in Data Forgetting for Sustainable Data Management	5472
<i>Ramon Rico, Arno Siebes, Yannis Velegarakis</i>	
Property Graph Standards: State of the Art & Open Challenges	5477
<i>Haridimos Kondylakis, Stefania Dumbrava, Matteo Lissandrini, Nikolay Yakovets, Angela Bonifati, Vasilis Efthymiou, George Fletcher, Dimitris Plexousakis, Riccardo Tommasini, Georgia Troullinou, Elisjana Ymeralli</i>	
Learned Cost Models for Query Optimization: From Batch to Streaming Systems	5482
<i>Roman Heinrich, Xiao Li, Manisha Luthra, Zoi Kaoudi</i>	
Filtered Vector Search: State-of-the-art and Research Challenges	5488
<i>Helena Caminal, Yannis Chronis, Yannis Papakonstantinou, Fatma Özcan, Anastasia Ailamaki</i>	
ML-Asset Management: Curation, Discovery, and Utilization	5493
<i>Mengying Wang, Moming Duan, Yicong Huang, Chen Li, Bingsheng He, Yinghui Wu</i>	
Machine Learning for Graph Data Management and Query Processing.....	5499
<i>Hanchen Wang, Ying Zhang, Wenjie Zhang</i>	
Database Perspective on LLM Inference Systems.....	5504
<i>James Pan, Guoliang Li</i>	

Panels

Beyond Incrementalism: How to Change the World Through Data Systems Research.....	5508
---	------

Viktor Leis

Where Does Academic Database Research Go From Here?..... 5510
Eugene Wu, Raul Castro Fernandez

Open Science: A New Paradigm for the Research Lifecycle..... 5512
Yannis Ioannidis

Panel on Neural Relational Data: Tabular Foundation Models, LLMs... or both? 5513
Paolo Papotti, Carsten Binnig

Endowment Awards

Versatile Property Graph Transformations 5516
Angela Bonifati

Disaggregation: A New Architecture for Cloud Databases..... 5527
Xiangyao Yu

Still Asking: How Good Are Query Optimizers, Really? 5531
Viktor Leis, Andrey Gubichev, Atanas Mirchev, Peter Boncz, Alfons Kemper, Thomas Neumann

Keynotes

Alphabets, Grammars, Calculators, and the End of Hand-Crafted Systems 5537
Stratos Idreos

Bridging Disciplines in Data Management Research to Solve Complex Data Problems..... 5538
Juliana Freire

Bringing the Operational and Analytical Worlds Together with Lakebase 5539
Matei Zaharia

TABLE OF CONTENTS

Research Papers

E2ETune: End-to-End Knob Tuning via Fine-tuned Generative Language Model	5540
<i>Xinmei Huang, Haoyang Li, Jing Zhang, Xinxin Zhao, Zhiming Yao, Yiyang Li, Tieying Zhang, Jianjun Chen, Hong Chen, Cuiping Li</i>	
SunStorm: Geographically distributed transactions over Aurora-style systems.....	5555
<i>Cuong Nguyen, Pooja Nilangekar, Heikki Linnakangas, Daniel Abadi</i>	
Sampling-based Predictive Database Buffer Management.....	5569
<i>Theo Vanderkooy, Mohammad Khalaji, Runsheng Benson Guo, Khuzaima Daudjee</i>	
AXE: A Task Decomposition Approach to Learned LSM Tuning	5582
<i>Andy Huynh, Anwesha Saha, Harshal Chaudhari, Manos Athanassoulis</i>	
Mayura: Exploiting Similarities in Motifs for Temporal Co-Mining	5596
<i>Sanjay Sri Vallabh Singapuram, Ronald Dreslinski, Nishil Talati</i>	
Mix & Match: Subgraph Matching for Absolute Coverage.....	5610
<i>Konstantinos Skitsas, Yuya Sasaki, Davide Mottin, Panagiotis Karras</i>	
In-depth Analysis of Graph-based RAG in a Unified Framework.....	5623
<i>Yingli Zhou, Yaodong Su, Youran Sun, Shu Wang, Taotao Wang, Runyuan He, Yongwei Zhang, Sicong Liang, Xilin Liu, Yuchi Ma, Yixiang Fang</i>	
CloudGlide: Deconstructing the Landscape of Cloud-Based Analytics.....	5638
<i>Michail Georgoulakis Misegiannis, Daniel Ritter, Viktor Leis, Jana Giceva</i>	
Algorithmic Data Minimization for Machine Learning over Internet-of-Things Data Streams	5652
<i>Ted Shaowang, Shinan Liu, Jonatas Marques, Nick Feamster, Sanjay Krishnan</i>	
LiquidCache: Efficient Pushdown Caching for Cloud-Native Data Analytics	5662
<i>Xiangpeng Hao, Andrew Lamb, Yibo Wu, Andrea Arpaci-Dusseau, Remzi Arpaci-Dusseau</i>	
Cleaning both Data Errors and Inaccurate Constraints on Numerical Sequential Data	5676
<i>Xiaoou Ding, Muyun Zhou, Yida Liu, Chen Wang, Hongzhi Wang, Jianmin Wang</i>	
Analyzing Near-Network Hardware Acceleration with Co-Processing on DPUs.....	5689
<i>Dimitrios Giouroukis, Dwi Nugroho, Varun Pandey, Steffen Zeuch, Volker Markl</i>	
An Evaluation of N-Gram Selection Strategies for Regular Expression Indexing in Contemporary Text Analysis Tasks	5703
<i>Ling Zhang, Shaleen Deep, Jignesh Patel, Karthikeyan Sankaralingam</i>	
Graph Transformers for Query Plan Representation: Potentials and Challenges.....	5716
<i>Chenghao Lyu, Guillaume Lachaud, Gabriel Lozano, Yanlei Diao</i>	

Exploring Exploratory Querying.....	5731
<i>Marcelo Arenas, Enrico Franconi, Janik Hammerer, Olaf Hartig, Katja Hose, Laura Koesten, George Konstantinidis, Leonid Libkin, Wim Martens, Yuya Sasaki, Stefanie Scherzinger, Katherine Thornton, Hsiang-Yun Wu</i>	
Finding Time-Proximity Communities in Temporal Heterogeneous Information Networks.....	5740
<i>Yifu Tang, Chengfei Liu, Lu Chen, Rui Zhou, Jianxin Li</i>	
Errata for "CGgraph: An Ultra-Fast Graph Processing System on Modern Commodity CPU-GPU Co-processor"	5753
<i>Pengjie Cui, Haotian Liu, Bo Tang, Ye Yuan</i>	