2026 IEEE 42nd International Conference on Data Engineering (ICDE 2026)

Montreal, QC, Canada
4-8 May 2026

Pages 1–728

IEEE Catalog Number: CFP26026-POD
ISBN: 979-8-3315-8366-8

Copyright © 2026, IEEE

All Rights Reserved

Copyright and Reprint Permissions:

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

*** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.

IEEE Catalog Number:	CFP26026-POD
ISBN (Print-On-Demand):	979-8-3315-8366-8
ISBN (Online):	979-8-3315-8365-1
ISSN:	1063-6382

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
E-mail:	curran@proceedings.com
Web:	www.proceedings.com



TABLE OF CONTENTS

Querying Historical k-Dense Subgraphs on Temporal Graphs 1
Qi Zhang, Yalong Zhang, Rong-Hua Li, Xu-Cheng Yin, Guoren Wang

On Graph Rewiring with Motifs: A Find-and-Replace Approach 14
Qihao Wang, Hongtai Cao, Xiaodong Li, Matin Najafi, Kevin Chen-Chuan Chang, Reynold Cheng

Compressing High-Frequency Time Series Through Multiple Models and Stealing From Residuals 28
Abduvoris Abduvakhobov, Søren Kejser Jensen, Christian Thomsen, Torben Bach Pedersen

Community-Level Personalized Recommendation by Exploiting Evolving User-Item Micro-Clusters 42
Xinyu Liu, Jinxia Guo, Qirui Hao, Zhongjing Yu, Qinli Yang, Junming Shao

RL-Paxos: Relieving the Leader's Burden with Efficient Task Offloading in Distributed Consensus 57
Chenhao Zhang, Jinquan Wang, Meng Han, Bing Wei, Xiaojian Liao, Limin Xiao, Shanchen Pang

AGRAG: Advanced Graph-Based Retrieval-Augmented Generation for LLMs 71
Yubo Wang, Haoyang Li, Fei Teng, Lei Chen

Effective Dataset Distillation for Spatio-Temporal Forecasting with BI-Dimensional Compression 86
Taehyung Kwon, Yeonje Choi, Yeongho Kim, Kijung Shin

Improving GPU Tensor Query Processing for Resource-Constrained Environments 100
Qian Xu, Feng Zhang, Shijie Gao, Kun Chen, Jianhua Wang, Zheng Chen, Xiaoyong Du

DistVec: Efficient Distributed Machine Learning in Parallel Database Systems 115
Xinyi Zhang, Liangzu Liu, Xupeng Miao, Yinjun Wu, Zhen Chen, Wei Lu, Xiaoyong Du, Bin Cu

RISK: Efficiently Processing Rich Spatial-Keyword Queries on Encrypted Geo-Textual Data 129
Zhen Lv, Cong Cao, Hongwei Huo, Jiangtao Cui, Yanguo Peng, Hui Li, Yingfan Liu

Boosting Small Language Models for Text-to-SQL with Fine-Grained Execution Feedback and Cost-Efficient Rewards 143
Thanh Dat Hoang, Thanh Trung Huynh, Matthias Weidlich, Thanh Tam Nguyen, Tong Chen, Hongzhi Yin Quoc Viet Hung Nguyen

A Set-Theoretic Approach to Detecting Logic Bugs in DBMS Inner Join Optimizations 157
Ce Lyu, Changzheng Wei, Yanhao Wang, Jie Liang, Li Lin, Hanghang Wu, Minghao Zhao, Ying Yan Aoying Zhou

ARCADE: A Real-Time Data System for Hybrid and Continuous Query Processing Across Diverse Data Modalities 171
Jingyi Yang, Songsong Mo, Jiachen Shi, Zihao Yu, Kunhao Shi, Xuchen Ding, Gao Cong

StreamShield: A Production-Proven Resiliency Solution for Apache Flink at Bytedance 185
Yong Fang, Yuxing Han, Meng Wang, Yifan Zhang, Yue Ma, Chi Zhang

FaScalSQL: A Fast and Scalable GPU-Accelerated SQL Query Engine for Out-of-Memory Tables 199
Chaemin Lim, Suhyun Lee, Jinwoo Choi, Kwanghyun Park, Jinho Lee, Joonsung Kim, Youngsok Kim

GeminiSketch: An Accurate and Efficient Sketch for Summarizing Temporal Graph Streams with Rolling-Out Elimination 215
Xuyang Jing, Chenhao Zhang, Zheng Yan, Qingze Jiang, Witold Pedrycz, Mingjun Wang, Cong Wang

Fast Discovery of Functional Dependencies via Bayesian Network Learning 229
Siyi Yang, Shenglin Chen, Xi Wang, Yuhua Tang, Ruochun Jin

Approximate Butterfly Counting in Sublinear Time 243
Chi Luo, Jiaxin Song, Yuhao Zhang, Kai Wang, Zhixing He, Kuan Yang

Exqutor: Extended Query Optimizer for Vector-Augmented Analytical Queries 257
Hyunjoon Kim, Chaerim Lim, Hyeonjun An, Rathijit Sen, Kwanghyun Park

RoarChain: A Robust Sharding Blockchain System for Enterprise Consortium 272
Yuan Sui, Xiaochun Yang, Bin Wang, Yujie Zhang, Lina Wang

Towards the Distributed Large-Scale k-NN Graph Construction by Graph Merge 285
Cheng Zhang, Wan-Lei Zhao, Shihai Xiao, Jiajie Yao, Xuecang Zhang

CSD-CoKV: Host-CSD Collaborative Offloading for High-Performance LSM-Tree Based KV Stores 298
Zhining Cao, Kai Zhang, Jinrun Yang, Hui Li, Nan Su, Qian Wei, Shikun Ma, Zehao Chen, Junbo Yin Haijun Zhang, Zhaoyan Shen

Conflict Resolution for Improving ML Accuracy 312
Wenfei Fan, Xiaoyu Han, Hufsa Khan, Weilong Ren, Yaoshu Wang, Min Xie, Zihuan Xu

MVGPT: Generative Materialized View Forecasting 326
Yue Han, Guoliang Li, Wenchun Xu, Xianglei Ran, Zeya Gong, Wei Guo, Guang Qiu, Bo Zheng

Accurate Table Question Answering with Accessible LLMs 340
Yangfan Jiang, Fei Wei, Ergute Bao, Yaliang Li, Bolin Ding, Yin Yang, Xiaokui Xiao

Autohformer: Efficient Hierarchical Autoregressive Transformer for Time Series Prediction 354
Qianru Zhang, Honggang Wen, Ming Li, Dong Huang, Siu-Ming Yiu, Christian S. Jensen, Pietro Liò

CausalPre: Scalable and Effective Data Pre-Processing for Causal Fairness 368
Ying Zheng, Yangfan Jiang, Kian-Lee Tan

Evolving Sketch: Time-Decaying Frequency Estimation for Evolving Streams 382
Ge Gao, Yang Du, He Huang, Yu-E Sun, Jianzhi Tang

Contextual Pattern Mining and Counting 395
Ling Li, Daniel Gibney, Sharma V. Thankachan, Solon P. Pissis, Grigorios Loukides

A Unified Framework for Compressed and Encrypted Text Direct Processing 409
Yani Liu, Feng Zhang, Yu Zhang, Siqi Ma, Elisa Bertino, Xiaoyong Du

Spatiotemporal Sketch Disaggregation: Streaming Analytics with Heterogeneous Resources 423
Jonatan Langlet, Peiqing Chen, Michael Mitzenmacher, Zaoxing Liu, Ran Ben Basat, Gianni Antichi

L4g: Two-Hop Label Management for Group Steiner Tree Search on Graphs 437
Xiaoyao Feng, Yahui Sun, Zhuoran Wang, Junlin Li, Sijia Luo, Rong-Hua Li

Revisiting Locally Differentially Private Protocols: Towards Better Trade-Offs in Privacy, Utility, and Attack Resistance 451
Héber H. Arcolezi, Sébastien Gambs

Micro: a Lightweight Middleware for Optimizing Cross-Store Cross-Model Graph-Relation Joins 465
Xiuwen Zheng, Arun Kumar, Amarnath Gupta

SwitchΔ: Asynchronous Metadata Updating for Distributed Storage with in-Network Data Visibility 479
Junru Li, Qing Wang, Zhe Yang, Shuo Liu, Jiwu Shu, Youyou Lu

DIndex: an Efficient on-Disk Learned Index for Memory-Constrained Environments 493
Jiahuan Shen, Chuzhe Tang, Haoning Lan, Ren Ren, Zhaoguo Wang

Federated Retrieval Over Embedding-Heterogeneous Vector Databases 507
Yuxiang Wang, Yongxin Tong, Zimu Zhou, Ziyuan He, Ruixi Hu, Ke Xu

LogDelta: Differential Encoding for Log Data 521
Songze Li, Shaoxu Song, Zhitao Shen

Secure Query Processing with Linear Online Cost 535
Qiyao Luo, Yilei Wang, Wei Dong, Ke Yi

VisPoison: An Effective Backdoor Attack Framework for Tabular Data Visualization Models 548
Shuaimin Li, Chen Jason Zhang, Xuanang Chen, Anni Peng, Zhuoyue Wan, Yuanfeng Song, Shiwen Ni Min Yang, Fei Hao, Raymond Chi-Wing Wong

PRO-HNSW: Proactive Repair and Optimization for High-Performance Dynamic HNSW Indexes 562
Huijun Jin, Jieun Lee, Shengmin Piao, Sangmin Seo, Sanghyun Park

Promoting Fairness in Information Access Within Social Networks 576
Changan Liu, Xiaotian Zhou, Ahad N. Zehmakan, Zhongzhi Zhang

Keyword-Aware Skyline Community Search on Semantics and Structure 590
Chuanhou Sun, Yuhai Zhao, Ling Li, Yuan Li

B^{S}-Tree: A Gapped Data-Parallel B-Tree 604
Dimitrios Tsitsigkos, Achilleas Michalopoulos, Nikos Mamoulis, Manolis Terrovitis

Zero-Knowledge Verifiable Graph Query Evaluation via Expansion-Centric Operator Decomposition 618
Hao Wu, Changzheng Wei, Yanhao Wang, Li Lin, Yilong Leng, Shiyu He, Minghao Zhao, Hanghang Wu Ying Yan, Aoying Zhou

cuFHEDB: GPU-Accelerated Fully Homomorphic Encryption Database 631
Shijie Gao, Feng Zhang, Qian Xu, Yang Li, XueFeng Liu, Chao Jiang, Limin Xiao, Siqi Ma, Elisa Bertino Xiaoyong Du

Contemp: Instance Caching Based on Container Temperature in Serverless Environment 646
Pengwei Wang, Nuo Chen, Haoquan Qi, Yichen Zhong, Shun Song

TopFGL: A Topology-Aware and Distributionagnostic Federated Learning Framework Tackling Topological Heterogeneity on Graph Data 660
Junyang Wang, Lan Zhang, Yihang Cheng, Mu Yuan, Tianfu Wang, Zhihui Fu, Jun Wang

OsmT: Bridging Openstreetmap Queries and Natural Language With Open-Source Tag-Aware Language Models 674
Zhuoyue Wan, Wentao Hu, Chen Jason Zhang, Yuanfeng Song, Shuaimin Li, Ruiqiang Xiao, Xiao-Yong Wei Raymond Chi-Wing Wong

Knapsack Optimization-Based Schema Linking for LLM-Based Text-to-SQL Generation 688
Zheng Yuan, Hao Chen, Zijin Hong, Qinggang Zhang, Feiran Huang, Qing Li, Xiao Huang

FLASH Viterbi: Fast and Adaptive Viterbi Decoding for Modern Data Systems 702
Ziheng Deng, Xue Liu, Jiantong Jiang, Yankai Li, Qingxu Deng, Xiaochun Yang

Sorting Compressed Time Series 716
Zhiheng Liu, Xingyu Liu, Shaoxu Song, Jianmin Wang

Reverse K Nearest Neighbor Query in Large Road Networks: a Tree Decomposition Based Approach 729
Dian Ouyang, Boyu Zhang, Jianye Yang, Shiyu Yang, Chonghua Wang, Xuemin Lin

Benchmarking RL-Enhanced Spatial Indices Against Traditional, Advanced, and Learned Counterparts 743
Guanli Liu, Renata Borovica-Gajic, Hai Lan, Zhifeng Bao

ProGQL: A Provenance Graph Query System for Cyber Attack Investigation 757
Fei Shao, Jia Zou, Zhichao Cao, Xusheng Xiao

Efficient Hypergraph Pattern Matching via Match-and-Filter and Intersection Constraint 772
Siwoo Song, Wonseok Shin, Kunsoo Park, Giuseppe F. Italiano, Zhengyi Yang, Wenjie Zhang

Trading Vector Data in Vector Databases 786
Jin Cheng, Xiangxiang Dai, Ningning Ding, John C.S. Lui, Jianwei Huang

Updatable Balanced Index for Fast on-Device Search with Auto-Selection Model 801
Yushuai Ji, Sheng Wang, Zhiyu Chen, Yuan Sun, Zhiyong Peng

Banknote-Chain: Achieving User-Incentivized Parallelism in Blockchain via a Banknote-Inspired Transaction Model 814
Ma Zhiyu, Li Xiaofeng, Zhao He, Zhou Tong, Sheng Nianzu, Cheng Haotian

Efficient Zero-Shot and Label-free Log Anomaly Detection for Resource-Constrained Systems 828
Zuohan Wu, Jiachuan Wang, Libin Zheng, Yongqi Zhang, Shuangyin Li, Lei Chen

Energy-Efficient Autonomous Driving With Adaptive Perception and Robust Decision 842
Yuyang Xia, Zibo Liang, Liwei Deng, Yan Zhao, Han Su, Kai Zheng

Mitigating Dual Load Imbalance via Dynamic Cooperative Scheduling in Distributed Key-Value Stores 856
Jiakun Zhang, Patrick P. C. Lee, Wenzhe Zhu, Yongkun Li, Shuyi Zhang, Yinlong Xu

MOCHI: Motif-Based Community Search Over Large Heterogeneous Information Networks 870
Yuhan Zhou, Qing Liu, Xin Huang, Jianliang Xu, Yunjun Gao

Reconfiguring Scalable Hashing with Persistent CPU Caches 884
Zhenyu Yu, Bolong Zheng, Ling Xu, Qianlu Wu, Qiang Chen, Ziyang Yue

TemplateQO: Template-Aware and Scalable Query Optimization with Data-Efficient Learning 897
Pengfei Zheng, Guoneng Li, Ling Xu, Rong Zhu, Yan Li, Bolong Zheng

Query-Driven Data Exploration with Heterogeneous Treatment Effects 910
Antonis Mandamadiotis, Sihem Amer-Yahia, Georgia Koutrika

SOLAR: Scalable Distributed Spatial Joins Through Learning-Based Optimization 923
Yongyi Liu, Ahmed Abdelmaguid, Ahmed R. Mahmood, Amr Magdy, Minyao Zhu

Interpreting Graph Inference with Skyline Explanations 937
Dazhuo Qiu, Haolai Che, Arijit Khan, Yinghui Wu

Chase Anonymisation: Privacy-Preserving Knowledge Graphs with Logical Reasoning 951
Luigi Bellomarini, Costanza Catalano, Andrea Coletta, Michela Iezzi, Pierangela Samarati

SSC-Join: An Efficient Syntactic-Semantic Collaboration Based Set Semantic Similarity Join Algorithm 965
Lianyin Jia, Chengchen Zeng, Mengjuan Li, Suprio Ray, Yinong Chen, Jiaman Ding, Xiuxing Li

Deferred Flushing for Out-of-Order Arrivals in Apache IoTDB 978
Xiaojian Zhang, Zhiheng Liu, Shaoxu Song, Xiangdong Huang, Chen Wang, Jianmin Wang

Efficient Traffic Forecasting on Large-Scale Road Network by Regularized Adaptive Graph Convolution 992
Kaiqi Wu, Weiyang Kong, Sen Zhang, Zitong Chen, Yubao Liu

SaSPartitioner: A Self-Adaptive Streaming Partitioner Using Deep Reinforcement Learning 1006
Shenghao Gong, Liu Liu, Ziquan Fang, Yunjun Gao, Yaofeng Tu

AdaFedRec: Adaptive Heterogeneous Federated Recommender Systems Across Multi-Device Users 1020
Zhenkai Li, Ming Hu, Chentao Jia, Yining Sun, Zhufeng Lu, Mingyang Yu, Yanxin Yang, Xiaofei Xie Mingsong Chen

MatKV: Trading Compute for Flash Storage in LLM Inference 1034
Kun-Woo Shin, Jay H. Park, Moonwook Oh, Yohan Jo, Jaeyoung Do, Sang-Won Lee

LUCID: An Updatable and Concurrent Learned Index for Larger-Than-Memory Data Management 1048
Chaohong Ma, Xiaohui Yu, Yifan Li, Aishan Maoliniyazi, Xiaofeng Meng

CoLSE: A Lightweight and Robust Hybrid Learned Model for Single-Table Cardinality Estimation Using Joint CDF 1062
Lankadinee Rathuwadu, Guanli Liu, Christopher Leckie, Renata Borovica-Gajic

An Encode-then-Decompose Approach to Unsupervised Time Series Anomaly Detection on Contaminated Training Data 1076
Buang Zhang, Tung Kieu, Xiangfei Qiu, Chenjuan Guo, Jilin Hu, Aoying Zhou, Christian S. Jensen Bin Yang

Grace: Alleviating Reconstruction Cost in Dynamic Graph Processing Systems 1091
Hongru Gao, Shuhao Zhang, Xiaofei Liao, Hai Jin

High-Fidelity Task Assignment in Spatial Crowdsourcing via Implicit Human Feedback 1105
Qingshun Wu, Yafei Li, Lei Gao, Guanglei Zhu, Lei Chen, Mingliang Xu

LAMP: A Dual-Mode Framework for Database Workload Memory Prediction 1119
Guoze Xue, Lu Chen, Ziquan Fang, Yushuai Li, Tianyi Li, Torben Bach Pedersen

Efficient Model-Agnostic Continual Learning for Next POI Recommendation 1132
Chenhao Wang, Shanshan Feng, Lisi Chen, Fan Li, Shuo Shang

GeoLayer: Towards Low-Latency and Cost-Efficient Geo-Distributed Graph Stores with Layered Graph 1146
Feng Yao, Xiaokang Yang, Shufeng Gong, Song Yu, Yanfeng Zhang, Ge Yu

Robust Index Benefit Estimation via Hierarchical and Two-Dimensional Feature Representation 1160
Tao Li, Feng Liang, Jinqi Quan, Zihang Yang, Teng Wang, Runhuai Huang, Xiping Hu, Meng Li Haipeng Dai

MTC: Scalable Transaction Commit for Multi-Primary Cloud Databases 1173
Kecheng Luo, Xiaoxian Wei, Wenxin Liu, Peng Cai, Aoying Zhou, Hui Li, Le Cai

EC-RAG: Towards Efficient Edge-Cloud Retrieval-Augmented Generation Systems 1186
Liang Wang, Kai Wang, Ranjun Jia, Kai Lu, Jiguang Wan, Hao Huo, Yulong Zhai, Zhiyuan Liang, Di Wang

TabLoft: Tabular Data Generation Based on LLM with Ordered Features 1201
Luyu Chen, Changhao Wu, Jingyi Li, Sen Liu, Guangnan Ye, Hongfeng Chai

CYANSQL: Unlock the Power of NL2SQL Via Clustering-Based Test-Time Scaling 1214
Haoyu Qin, Tonghui Ren, Zhenying He, X.Sean Wang, Jiashu Xing, Yanghuan Ye, Shifei Huang, Jinbao Li

Effective Fairest Community Search Over Heterogeneous Information Networks 1230
Taige Zhao, Jianxin Li, Man Li, Wei Luo, Jingxian Cheng, Yuan Miao, Hua Wang

Iit-Tree: an Efficient Index to Support Interval-Based Query on Large Temporal Graphs 1243
Faming Li, Shengli Qiu, Xiaochun Yang, Bin Wang, Hengzhao Ma

PORCA: Root Cause Analysis with Partially Observed Data 1256
Chang Gong, Di Yao, Jin Wang, Wenbin Li, Lanting Fang, Yongtao Xie, Kaiyu Feng, Peng Han, Jingping Bi

Efficient and Scalable Search for Statistics 1270
Antoine Gauquier, Simon Ebel, Helena Galhardas, Théo Galizzi, Ioana Manolescu, Aurélien Peden Pierre Senellart

Reconstructing TensorLog for Scalable End-to-End Rule Learning 1284
Kunxun Qi, Jianfeng Du, Hai Wan, Wei Wang

HL-Index: Fast Reachability Query in Hypergraphs 1298
Peiting Xie, Xiangjun Zai, Yanping Wu, Xiaoyang Wang, Wenjie Zhang, Lu Qin

Astraea: Efficient Pipelined Micro-Batch Stream Processing with Non-Hash Differentiated Partitioning 1312
Sijie Wu, Hanhua Chen, Hai Jin, Haoran Cai

Trajectory–User Linking via Heterogeneous Preference Graph and Dual-Encoder Mutual Distillation 1325
Zeming Tian, Zixin Qin, Huaijie Zhu, Ningning Cui, Wei Liu, Jianxing Yu, Jian Yin

PC-PS: A Multi-Dimensional Point-Cloud Data Publish/Subscribe System 1338
Yuanchi Fan, Lisi Chen, Shuo Shang, Christian S. Jensen

TransLGX: A Self-Contained Model to Predict the Entire Lifecycle and Complete State of Logistics Package Trajectories 1352
Yichen Song, Jianfeng Zhou, Jian-Ya Ding, Renhao Cao

SHMemora: Protective Key-Value Store on Distributed Shared Memory 1366
Jiajun Luo, Siyu Lin, Yunpeng Xu, Shengwei Liu, Jin Xia, Dong Liu, Zheng Liu, Huanchen Zhang, Teng Ma Shuwen Deng

More Than Pivot for Maximal Clique Enumeration 1380
Zhaoyi Zhong, Rui Zhou, Lu Chen, Xiaofan Li, Chengfei Liu

Provsql: a General System for Keeping Track of the Provenance and Probability of Data 1394
Aryak Sen, Silviu Maniu, Pierre Senellart

Data Guard: A Fine-Grained Purpose-Based Access Control System for Large Data Warehouses 1408
Khai Tran, Sudarshan Vasudevan, Pratham Desai, Alex Gorelik, Mayank Ahuja
Athrey Yadatore Venkateshababu, Mohit Verma, Dichao Hu, Walaa Eldin Moustafa, Vasanth Rajamani
Ankit Gupta, Issac Buenrostro, Kalinda Raina, Yanwen Lin

Fine-Grained Manipulation Attacks to Local Differential Privacy Protocols for Range Query 1421
Xinyu Li, Wenda Chen, Xuebin Ren

GPU-Accelerated OLTP: An in-Depth Analysis of Concurrency Control Schemes 1434
Zihan Sun, Yuyu Luo, Yong Zhang, Chao Li, Chunxiao Xing

Semantic Publish/Subscribe Over Evolving Topics 1448
Yiming Yao, Lisi Chen, Shuo Shang

FusAD: Time-Frequency Fusion with Adaptive Denoising for General Time Series Analysis 1462
Da Zhang, Bingyu Li, Zhiyuan Zhao, Feiping Nie, Junyu Gao, Xuelong Li

Damba-ST: Domain-Adaptive Mamba for Efficient Urban Spatio-Temporal Prediction 1477
Rui An, Yifeng Zhang, Ziran Liang, Wenqi Fan, Yuxuan Liang, Xuequn Shang, Qing Li

Beyond Traditional Diagnostics: Transforming Patient-Side Information Into Predictive Insights with Knowledge Graphs and Prototypes 1491
Yibowen Zhao, Yinan Zhang, Zhixiang Su, Lizhen Cui, Chunyan Miao

SSFusion: Tensor Fusion with Selective Sparsification for Efficient Distributed DNN Training 1505
Zhangqiang Ming, Rui Wang, Yuchong Hu, Yuanhao Shu, Wenxiang Zhou, Xinjue Zheng, Dan Feng

MojoFrame: Dataframe Library in Mojo Language 1519
Shengya Huang, Zhaoheng Li, Derek Werner, Yongjoo Park

Text2VectorSQL: Towards a Unified Interface for Vector Search and SQL Queries 1534
Zhengren Wang, Dongwen Yao, Bozhou Li, Dongsheng Ma, Bo Li, Zhiyu Li, Feiyu Xiong, Bin Cui
Linpeng Tang, Wentao Zhang

Rethinking Flexible Graph Similarity Computation: One-Step Alignment with Global Guidance 1548
Zhouyang Liu, Ning Liu, Yixin Chen, Jiezhong He, Shuai Ma, Dongsheng Li

Beyond Homophily: Community Search on Heterophilic Graphs 1562
Qing Sima, Xiaoyang Wang, Wenjie Zhang

Efficient Cloud-Edge Collaborative Approaches to Sparql Queries Over Large RDF Graphs 1576
Shidan Ma, Peng Peng, Xu Zhou, M. Tamer Özsu, Lei Zou, Guo Chen

Generalizable Address-Aware Semantic Prefetching for Scalable Transactional and Analytical Workloads 1590
Farzaneh Zirak, Farhana Choudhury, Renata Borovica-Gajic

AlignSketch: A Framework for Aligning Theoretical and Practical Estimation Errors 1604
Ce Zheng, Hanyue Zheng, Jingwei Shi, Xinye Xu, Wei Zhou, Tong Yang, Zhenyu Guan, Yong Cui

FedCurrMM: A Federated Map Matching Framework with Curriculum-Aware Client Selection 1618
Minxiao Chen, Haitao Yuan, Haoning Wang, Nan Jiang, Zhihan Zheng, Ao Zhou, Shangguang Wang

Enabling Homomorphic Analytical Operations on Compressed Scientific Data with Multi-Stage Decompression 1632
Xuan Wu, Sheng Di, Tripti Agarwal, Kai Zhao, Xin Liang, Franck Cappello

DNA: A Distribution-and-Aggregation Solution for Spatiotemporal K-Function-Based Analysis 1646
Tsz Nam Chan, Bojian Zhu, Dingming Wu, Renchi Yang, Ruisheng Wang

Chubby: Robust Smart Contract Execution Against Dependency Over-Declaration 1660
Junyu Wei, Xiaodong Qi, Qifeng Que, Zhao Zhang, Yanqin Yang, Cheqing Jin

PLAN: Fast and Approximate Gaussian Kernel Density Visualization in Road Networks 1674
Tsz Nam Chan, Hongwei Ye, Bojian Zhu, Leong Hou U, Dingming Wu, Ruisheng Wang Joshua Zhexue Huang

MINOR: Multivariate Time Series Iterative Cleaning Algorithm 1689
Aoqian Zhang, Yinru Sun, Pengxiang Hao, Yifeng Gong, Boyang Li, Jing Geng, Zheng Wang Lianpeng Qiao

From Single to Multiple Attributes: Experimental Insights on Sampling-Based Distinct Combination Estimation in Group-by Queries 1702
Yujie Zhang, Xiaochun Yang, Bin Wang, Yuan Sui

Explaining GNN Negatives Globally and Locally 1716
Kehan Pang, Wenfei Fan, Min Xie, Dandan Lin

Scaling Subsequence Similarity Join Based on Dynamic Time Warping 1730
Zemin Chao, Qiaoyi Zheng, Xingxing Xiao, Boyu Xiao, Zhixin Qi, Hongzhi Wang

Overcoming the Sync-Compute Dilemma in Parallel Graph-Based Vector Retrieval 1743
Qiji Mo, Zhiyuan Hua, Zebin Yao, Lixiao Cui, Gang Wang, Xiaoguang Liu, Zijing Wei, Xinyu Liu Tianxiao Tang, Shaozhi Liu, Lin Qu

Procore: Robust Core-Set Selection Via Pareto Multi-Dimensional Optimization From Noisy Data 1758
Xiaoou Ding, Hongbin Hu, Songnan Jiang, Muyun Zhou, Chen Wang, Jingru Yang, Hongzhi Wang

NebulaStream: An Adaptive and Efficient Multi-Query Stream Processing Engine 1772
Nils L. Schubert, Lukas Schwerdtfeger, Sara Schnaterbeck, Philipp M. Grulich, Bonaventura Del Monte Steffen Zeuch, Volker Markl

An LLM-Guided Query-Aware Inference System for GNN Models on Large Knowledge Graphs 1788
Waleed Afandi, Hussein Abdallah, Ashraf Aboulnaga, Essam Mansour

Tetris: Lightweight Hyperparameter Auto-Tuning for Mitigating Performance Spikes in LSM-KVS 1802
Yina Lv, Wenhao Zhu, Qiao Li, Quanqing Xu, Congming Gao, Chuanhui Yang, Xiaoli Wang Chun Jason Xue

An End-To-End Re-Evaluation of Table Entity-Linkers 1816
Martin Pekár Christensen, Matteo Lissandrini, Katja Hose

Cole ${ }^{+}$: Towards Practical Column-Based Learned Storage for Blockchain Systems 1830
Ce Zhang, Cheng Xu, Haibo Hu, Jianliang Xu

Efficient Meta-Path Constrained Reachability Query on Heterogeneous Information Networks 1844
Chao Ni, Zi Chen, Long Yuan, Bolong Zheng, Lu Qin

HEXGEN-FLOW: Optimizing LLM Inference Request Scheduling for Agentic Text-to-SQL 1858
You Peng, Youhe Jiang, Wenqi Jiang, Chen Wang, Binhang Yuan

Krone: Hierarchical and Modular Log Anomaly Detection 1872
Lei Ma, Jinyang Liu, Tieying Zhang, Peter M. VanNostrand, Dennis M. Hofmann, Lei Cao Elke A. Rundensteiner, Jianjun Chen

MINT: Multi-Vector Search Index Tuning 1887
Jiongli Zhu, Yue Wang, Bailu Ding, Philip A. Bernstein, Vivek Narasayya, Surajit Chaudhuri

LLM4Hint: Leveraging Large Language Models for Hint Recommendation in Offline Query Optimization 1902
Suchen Liu, Yang Lin, Yinjun Han, Jun Gao

CFDGraph: Privacy-Preserving Graph Processing for Large-Scale Collaborative Fraud Detection 1916
Qiulin Wu, Amelie Chi Zhou, Tristan Allard, Shadi Ibrahim, Yuhong Feng, Lichun Li, Amr El Abbadi

SNI-GNN: SmartNIC-Assisted Full-Graph GNN Training with In-Network Embedding Prediction 1930
Guofan Yu, Sitian Chen, Zhenheng Tang, Xiaowen Chu, Amelie Chi Zhou

A Robust and Globally-Accurate Hierarchical Hub Labeling Index for SP-Distance Queries in Dynamic Road Networks 1944
Wei Liu, Ziqiang Yu, Xiaohui Yu, Yang Liu, Simu Liu

TRADER: Real-time Arbitrage Detection via Negative Cycles on Dynamic Graphs 1957
Bingqiao Luo, Yuhang Chen, Jiaxin Jiang, Yuheng Cong, Ziyu He, Shixuan Sun, Bingsheng He Wee Howe Ang

Decomposition-Driven Multi-Table Retrieval and Reasoning for Numerical Question Answering 1971
Feng Luo, Hai Lan, Hui Luo, Zhifeng Bao, Xiaoli Wang, J.Shane Culpepper, Shazia Sadiq

HaS: Accelerating RAG Through Homology-Aware Speculative Retrieval 1985
Peng Peng, Weiwei Lin, Wentai Wu, Xinyang Wang, Yongheng Liu

SaCal: An Efficient Saliency-Guided Causal Framework for Interpretable Healthcare Analytics 1999
Feixuan Lin, Chenyu You, Zhongle Xie, Zhaojing Luo, Meihui Zhang

BOND: A Co-Designed Framework for LLM-Powered Analytics Over Relational Data 2013
Lixiang Chen, Qin Zheng, Zhicheng Pan, Chengcheng Yang, Rong Zhang, Xuan Zhou

Efficient Top-k Nearest Neighbors Search in Dynamic Road Networks 2026
Junhua Zhang, Yamei Song, Wentao Li, Lu Qin

An Efficient and Scalable Approach for Path Queries on Public Transportation Networks 2039
Junhua Zhang, Wentao Li, Wenjie Zhang, Lu Qin, Xiaochun Yang

Truth $\neq$ Frequency: Leveraging Dependencies for Subset Repair 2053
Haoda Li, Jiahui Chen, Yu Sun, Shaoxu Song, Haiwei Zhang, Xiaojie Yuan

iKSP: A Path Enumeration Index in Road Networks 2067
Zihan Luo, Lei Li, Mengxuan Zhang, Xinjie Zhou, Zizhuo Xu, Xiaofang Zhou

Novel Table Search 2080
Besat Kassaie, Renée J. Miller

Query-Guided Analysis and Mitigation of Data Verification Errors 2093
Ran Schreiber, Yael Amsterdamer

ImmortalChopper: Real-Time and Resilient Distributed Transactions in the Edge-Cloud 2107
Juncheng Fang, Farzad Habibi, Binbin Gu, Faisal Nawab

DIFFCOM: Conditional Discrete Diffusion Model for Community Search 2120
Ling Li, Liang Bai, Siqiang Luo, Yejiang Wang, Yuhai Zhao

GECO: A Confidentiality-Preserving and High-Performance Permissioned Blockchain Framework for General Smart Contracts 2134
Songxiao Guo, Rongxin Guan, Ji Qi, Zongyuan Zhang, Tianyang Duan, Sen Wang, Yanjun Wu, Heming Cui

Distance Comparison Operations are not Silver Bullets in Vector Similarity Search: A Benchmark Study on their Merits and Limits 2148
Zhuanglin Zheng, Yuxiang Zeng, Chenchen Liu, Yunzhen Chi, Binhan Yang, Yongxin Tong

Efficient Community Search on Attributed Public-Private Graphs 2162
Yuqi Chen, Weihan Zhang, Xin Huang

Batcher: Learning to Construct Cost-Efficient Batches of Small Queries in Big Data Processing Platforms 2176
Yeonsu Park, Taesung Lee, Byungchul Tak, Wook-Shin Han

UTune: Towards Uncertainty-Aware Online Index Tuning 2191
Chenning Wu, Sifan Chen, Wentao Wu, Yinan Jing, Zhenying He, Kai Zhang, X. Sean Wang

RAMSeS: Robust and Adaptive Model Selection for Time-Series Anomaly Detection Algorithms 2205
Mohamed Abdelmaksoud, Sheng Ding, Andrey Morozov, Ziawasch Abedjan

CARROT: A Learned Cost-Constrained Retrieval Optimization System for RAG 2220
Ziting Wang, Haitao Yuan, Wei Dong, Gao Cong, Feifei Li

SLGParser: Practical and Efficient Label-Free Log Parsing Using Large Language Models 2234
Yibing Hu, Cong Wang, Lixin Zhao, Aimin Yu

Unifying Graph Traversals and Time Series Joins in Hybrid Graphs 2248
Gianluca Rossi, Angela Bonifati, Riccardo Tommasini

RFOD: Random Forest-Based Outlier Detection for Mixed-Type Tabular Data 2262
Yihao Ang, Peicheng Yao, Yifan Bao, Yushuo Feng, Qiang Huang, Anthony K. H. Tung, Zhiyong Huang

CAMMSR: Category-Guided Attentive Mixture of Experts for Multimodal Sequential Recommendation 2276
Jinfeng Xu, Zheyu Chen, Shuo Yang, Jinze Li, Hewei Wang, Yijie Li, Jianheng Tang, Yunhuai Liu Edith C. H. Ngai

CactusDB: Unlock Co-Optimization Opportunities for SQL Queries and AI/ML Model Inferences 2290
Lixi Zhou, Kanchan Chowdhury, Lulu Xie, Jaykumar Tandel, Hong Guan, Zhiwei Fan, Xinwei Fu, Jia Zou

Time-Varying Vector Field Compression with Preserved Critical Point Trajectories 2305
Mingze Xia, Yuxiao Li, Pu Jiao, Bei Wang, Xin Liang, Hanqi Guo

Geography-Aware Large Language Models for Next POI Recommendation 2320
Wei Liu, Zhao Liu, Muzu Xie, Huaijie Zhu, Jianxing Yu, Jian Yin, Wang-Chien Lee

Label-Constrained Column Annotation with Language Models and Graph Neural Networks 2334
Duo Yang, Ioannis Dasoulas, Anastasia Dimou

When Complex Event Recognition Meets Cloud-Native Architectures 2350
Shizhe Liu, Haipeng Dai, Meng Li, Yuemeng Zhang, Shaoxu Song, Zhifeng Bao, Hancheng Wang Xiaofeng Gao, Guihai Chen

ABC: Numerical Data Collection Under Local Differential Privacy Without Prior Knowledge 2364
Incheol Baek, Hyungbin Kim, Yon Dohn Chung

REMON: Remote External Memory Over the Network 2377
Shiquan Zhang, Michail Bachras, Yuqiu Zhang, Yunhao Mao, Hans-Arno Jacobsen

Incremental GNN Embedding Computation on Streaming Graphs 2391
Qiange Wang, Haoran Lv, Yanfeng Zhang, Weng-Fai Wong, Bingsheng He

Lightweight 2-Hop Labels for Reachability Queries on Large-Scale Graphs 2405
Yishu Wang, Jinlong Chu, Ye Yuan, Yu Gu, Lianpeng Qiao

WikiDBGraph: A Data Management Benchmark Suite for Collaborative Learning Over Database Silos 2418
Zhaomin Wu, Ziyang Wang, Bingsheng He

Clue-RAG: Towards Accurate and Cost-Efficient Graph-Based RAG Via Multi-Partite Graph-Based Index 2432
Yaodong Su, Yixiang Fang, Yingli Zhou, Chuanhui Yang

VIREO: Human-in-the-Loop DBMS Fuzzing with Visualization and LLM Support 2447
Jie Liang, Zhiyong Wu, Jingzhou Fu, Chi Zhang, Runpei Miao, Zhuo Su, Yu Jiang, Shuai Ma

LEAF-SQL: Level-Wise Exploration with Adaptive Fine-Graining for Text-to-SQL Skeleton Prediction 2462
Zhao Tan, Xiping Liu, Qing Shu, Qizhi Wan, Dexi Liu, Changxuan Wan

An Efficient and Effective Evaluator for Text2SQL Models on Unseen and Unlabeled Data 2476
Trinh Pham, Thanh Tam Nguyen, Viet Huynh, Hongzhi Yin, Quoc Viet Hung Nguyen

PAT: Towards Transaction Routing with Page Affinity in Shared-Cache Databases 2490
Zhongqin Tan, Haoyuan Zhang, Yanfeng Zhang, Zeshun Peng, Weixing Zhou, Jinyu Zhang, Yang Ren Guoliang Li, Ge Yu

Density Decomposition of Multilayer Graphs 2504
Jiaqi Jiang, Rong-Hua Li, Yalong Zhang

AOEH: An Efficient Extendable Hashing to Reduce Read/Write Amplification for Persistent Memory 2517
Shihao Zhang, Chi Zhang, Yunfei Gu, Chentao Wu, Jie Li, Junzhe Lv

Nezha: A Key-Value Separated Distributed Store with Optimized Raft Integration 2531
Yangyang Wang, Yucong Dong, Ziqian Cheng, Zichen Xu

SpendableStore: A UTXO-Based Decentralized Data Store 2545
Yinan Zhou, Faisal Nawab

Process Faster, Pay Less: Functional Isolation for Stream Processing 2559
Eleni Zapridou, Michael Koepf, Panagiotis Sioulas, Ioannis Mytilinis, Anastasia Ailamaki

SQAC: Scalable Querying of Attribute-Constrained (α, β)-Cores over Large Bipartite Graphs 2573
Xin Deng, Peng Peng, Baoqing Sun, Shuo Dai, Zheng Qin, Lijun Chang

Listing Minimal Cores in Large Real-World Graphs 2587
Yukai Sun, Kaiqiang Yu, Shengxin Liu, Cheng Long, Raymond Chi-Wing Wong, Xun Zhou, Min Zhang

Time-Frequency Conditioned Diffusion for Multivariate Time Series Imputation 2601
Yumeng Liu, Zheng Wang, Jikui Liu, Kaisa Zhang, Weidong Gao, Xiaomao Fan

Low-Latency Stateful Stream Processing Through Timely and Accurate Prefetching 2615
Eleni Zapridou, Anastasia Ailamaki

Hydra: Breaking the Global Ordering Barrier in Multi-BFT Consensus 2629
Hanzheng Lyu, Shaokang Xie, Jianyu Niu, Mohammad Sadoghi, Yinqian Zhang, Cong Wang Ivan Beschastnikh, Chen Feng

Rangereduce: Query-Driven LSM Compactions 2643
Shubham Kaushik, Manos Athanassoulis, Subhadeep Sarkar

F5: A Robust SIMD-Accelerated MSD Radix Sort 2657
Arif Arman, Dmitri Loguinov

SQLMorph: Query Mutation and Fine-Grained Metrics for Text-to-SQL Evaluation 2670
Mohammadhossein Malekpour, Mohamed Riahi, Maxime Lamothe, Amine Mhedhbi

Telescope: A Learned What-If Call for Column Store Selection in HTAP Databases 2683
Yidong Zhang, Chao Zhang, Zhengkun Wu, Ju Fan, Xinyi Zhang, Hong Chen, Yuxing Chen, Anqun Pan

TORepair: Diffusion-Based Task-Oriented Error Repair Via Differentiable Bi-Level Optimization 2696
Wei Ni, Xiaoye Miao, Xiangyu Zhao, Yangyang Wu, Jianwei Yin

Table Question Answering via Adaptive Routing 2710
Yang Liu, Mengyi Yan, Jiao Xue, Weilong Ren, Yutong Ye, Haoyi Zhou, Jianxin Li, Zhumin Chen

[Experiment, Analysis, and Benchmark] BEACON: A Benchmark for Efficient and Accurate Counting of Subgraphs 2725
Xiangju Zhu, Mohammad Matin Najafi, Chrysanthi Kosyfaki, Xiaodong Li, Reynold Cheng Laks V.S. Lakshmanan

XRAG: Examining the Core - Benchmarking Foundational Components in Advanced Retrieval-Augmented Generation 2739
Qili Zhang, Qianren Mao, Yangyifei Luo, Yashuo Luo, Hanwen Hao, Zhilong Cao, Weifeng Jiang Zhijun Chen, Junnan Liu, Feng Yan, Xiaolong Wang, Jinlong Zhang, Zhenting Huang, Zhixing Tan, Jie Sun Bo Li, Jianxin Li, Philip S. Yu

A-Scan: Efficient Scale-Up Analytics via Throughput-Guided Data Movement 2753
Hamish Nicholson, Aunn Raza, Viktor Sanca, Anastasia Ailamaki

C²TC: A Training-Free Framework for Efficient Tabular Data Condensation 2767
Sijia Xu, Fan Li, Xiaoyang Wang, Zhengyi Yang, Xuemin Lin

ShareFlow: An Efficient Framework for Multi-Query Continuous Subgraph Matching 2781
Peiqi Yuan, Zhaohang Feng, Ruiqi Xu, Keming Li, Rui Mao, Bo Tang

L^{3} C: Leaf-Centric Continuous Codes for Natural Language-Driven Table Discovery 2794
Qiyuan Zhang, Ruochun Jin, Jixin Zhang, Yuhua Tang, Xiang Zhao, Shixuan Liu

APEX: Adaptive Variable-Wise Parallel Execution for Worst-Case Optimal Joins on Graph Queries 2808
Yipeng Liu, Yuming Lin, Zhicheng Pan, Chengcheng Yang, You Li, Aoying Zhou

HistCore: Scalable k-Core Decomposition on GPUs with Locality-Aware Computation 2822
Chen Zhao, Guojia Wan, Ting Yu, Jiawei Jiang, Bo Du

SINDI: An Efficient Index for Sparse Vector Approximate Maximum Inner Product Search 2835
Ruoxuan Li, Xiaoyao Zhong, Jiabao Jin, Peng Cheng, Wangze Ni, Zhitao Shen, Wei Jia, Xiangyu Wang Heng Tao Shen, Jingkuan Song

Information Leakage From Prices in Query-Based Data Markets 2849
Teng Tu, Huanhuan Peng, Xiaoye Miao, Guanjie Cheng, Shuiguang Deng, Jianwei Yin

Maximum Balanced Clique Search on Large Directed Graphs 2863
Jianhua Wang, Jianye Yang, Zhaoquan Gu, Dian Ouyang, Ziyi Ma, Ying Zhang

Online Multi-Modal Spatio-Temporal Prediction: a Reinforcement Learning and Dynamic Contrastive Framework 2877
Ziquan Fang, Tinghui Luo, Xiaole Pan, Lu Chen, Surun Ji, Mingfan Lu

Mirror Asymmetry Perfect Hashing: A Memory-Efficient and Load-Intensive-Optimized Hashing Index on Hybrid DRAM-PMem Architecture 2891
Jingcheng Ju, Zirui Liu, Feng Liu, Kaicheng Yang, Yikai Zhao, Tong Yang, Guodong Yang, Xingchun Wang Duohe Ma

ELENA: AN Explainability-Aided Online Query Optimization Framework 2905
Yuan Dong, Yuanyuan Yao, Yangyang Wu, Lu Chen, Rong Zhu

PRIME: Efficient Algorithm for Token Graph Routing Problem 2918
Haotian Xu, Yuqing Zhu, Yuming Huang, Jing Tang

Fast and Accurate Element-Level Streaming CP Decomposition for Higher-Order Tensors 2931
Jeongyoung Lee, SeungJoo Lee, U Kang

Answering Federated Range Queries with Local Differential Privacy 2945
Yuemin Zhang, Qingqing Ye, Junxu Liu, Wei Dong

GoCache: Accelerating Out-Of-Core Graph Queries with Pattern-Driven Caching 2958
Zheng Yang, Yicheng Zhang, Lixiao Cui, Luofan Chen, Chongzhuo Yang, Xiaojian Luo, Sijie Shen Wenyuan Yu, Jingren Zhou, Cheng Li

TS3D: A Temporal Multimodal Dataset for Distributed Database System Analysis 2972
Yuanyuan Yao, Yuhan Shi, Yian Wei, Lu Chen, Mourad Khayati, Cheng Long, Tianyi Li

SkyNet: Solving Skyline Queries with Neural Networks 2985
Jinfei Liu, Jiayao Zhang, Pengyun Zhu, Li Xiong, Jian Pei

Llmia: an Out-Of-The-Box Index Advisor Via in-Context Learning With Llms 2999
Xinxin Zhao, Xinmei Huang, Haoyang Li, Jing Zhang, Shuai Wang, Tieying Zhang, Jianjun Chen, Rui Shi Cuiping Li, Hong Chen

Approximate Diverse k-Nearest Neighbor Search in Vector Database 3013
Jiachen Zhao, Xiao Yan, Eric Lo

Balancing Competition for Fairness-Aware Task Recommendation and Assignment in Spatial Crowdsourcing 3026
Jinwen Chen, Hao Miao, Lei Jia, Guangqiang Yin, Yan Zhao, Kai Zheng

RaSE-KGC: A Relation-Aware Segment Encoding Approach for Knowledge Graph Completion 3039
Chenxiao Lin, Ye Luo, Kunhong Liu, Qingqiang Wu

C²graph: A Compression-Collaboration Algorithm for CPU-GPU Hybrid Weighted Graph Traversals 3053
Ning Wang, Huaibei Li, Shen Su, Yu Gu, Ge Yu, Zhigang Wang, Dawei Zhao, Hui Lu, Zhihong Tian

Accelerating Metadata Management of DFS via Speculative Permission Checking 3067
Yiduo Wang, Linghang Meng, Liang Li, Jie Wu

Prompt-Guided Community Search Under Extreme Few-Shot Supervision 3081
Wenxin Yang, Kaiyu Feng, Lanting Fang, Kangfei Zhao, Xia Wu

Robust Single-Message Shuffle Differential Privacy Protocol for Accurate Distribution Estimation 3095
Xiaoguang Li, Hanyi Wang, Yaowei Huang, Jungang Yang, Qingqing Ye, Haonan Yan, Ke Pan, Zhe Sun Hui Li

City-Wide Origin-Destination Matrix Generation via Cascaded Graph Denoising Diffusion 3109
Can Rong, Jingtao Ding, Zhicheng Liu, Peng Lu, Yong Li

Beyond Imputation: A Semantic Unification Framework for Data and its Missingness in Multimodal Healthcare Analytics 3123
Chaohe Zhang, Liantao Ma, Shiwei Lyu, Xin Gao, Junfeng Zhao, Yasha Wang, Xu Chu

VisiFold: Long-Term Traffic Forecasting via Temporal Folding Graph and Node Visibility 3137
Zhiwei Zhang, Xinyi Du, Weihao Wang, Xuanchi Guo, Wenjuan Han

Resystance: Unleashing Hidden Performance of Compaction in LSM-Trees Via eBPF 3152
Hongsu Byun, Seungjae Lee, Honghyeon Yoo, Myoungjoon Kim, Sungyong Park

EDDI: Explaining Data Drift Using Influence 3166
Nikolaos Myrtakis, Andrea Castellani, Ioannis Tsamardinos, Vassilis Christophides

Subtree Mode and Applications 3180
Jialong Zhou, Ben Bals, Matei Tinca, Ai Guan, Panagiotis Charalampopoulos, Grigorios Loukides Solon P. Pissis

HCT-QA: A Benchmark for Question Answering on Human-Centric Tables 3196
Mohammad S. Ahmad, Zan Naeem, Michaël Aupetit, Ahmed Elmagarmid, Mohamed Eltabakh, Xiaosong Ma Mourad Ouzzani, Chaoyi Ruan, Hani Al-Sayeh

MM2SQL: A Benchmark and Method for Visually-Grounded SQL Generation 3211
Shengze Shi, Tao Ren, Li Qi, Tingrui Yang, Wei Xiong, Jun Hu

Doux: Decoupling Values from Keys for Real-Time Analytics 3225
Shiming Yang, Yu Luo, Shuang Liu, Wei Lu, Kuien Liu, Yuxing Chen, Anqun Pan, Lixiong Zheng Xiaoyong Du

Analysis of Candidate Keys in Relational Databases 3238
Zihui Yang, Yuqian Ma, Sebastian Link

Lequa: A Learning-Based Query-Aware Framework for Selective Query Optimization 3252
Guoneng Li, Pengfei Zheng, Ling Xu, Yan Li, Bolong Zheng

Robust Spatial-Temporal Similar Trajectory Search via Structure-Enhanced Domain-Invariant Learning 3265
Xiaolin Han, Yonghao Zhou, Chenhao Ma, Lingyun Song, Xinbiao Gan, Xuequn Shang

LLMSQLMUTATOR: LLM-Powered Test Case Generation for Database Using Bug Reports 3279
Chenglin Tian, Chaofan Li, Yawen Li, Yingxia Shao

Toward Scalable Tucker Decomposition: Skew-Aware Multi-Level Partitioning with GPU–Storage Co-Processing 3293
Seung Hyeon Song, Jihye Lee, Chanki Kim, Kang-Wook Chon

Mitigating GenAI-Powered Evidence Pollution for Out-Of-Context Misinformation Detection 3307
Zehong Yan, Peng Qi, Wynne Hsu, Mong Li Lee

Efficient Query Rewrite Rule Discovery via Standardized Enumeration and Learning-to-Rank 3322
Yuan Zhang, Yuxing Chen, Yuekun Yu, Jinbin Huang, Rui Mao, Anqun Pan, Lixiong Zheng, Jianbin Qin

Text2sql-Flow: a Robust Sql-Aware Data Augmentation Framework for Text-To-Sql 3336
Qifeng Cai, Hao Liang, Chang Xu, Tao Xie, Wentao Zhang, Bin Cui

Efficient Size Constraint Community Search Over Heterogeneous Information Networks 3350
Xinjian Zhang, Chengfei Liu, Lu Chen, Rui Zhou, Bo Ning

SOLAR: Efficient Spatial Queries on Real-Time LSM-Based Storage 3363
Jingyi Yang, Jiachen Shi, Jian Chen, Gao Cong

SQLVec: SQL-Based Vector Similarity Search 3377
Zhequn Zhang, Yuanyuan Zhu, Hao Zhang, Jeffrey Xu Yu

Data-Segmentation Prompt Based Continual Learning Framework for Online Spatio-Temporal Prediction 3391
Banglie Yang, Liwei Deng, Cheng Dai, Kai Zheng

Efficient Graph Matching with Pattern Reduction 3405
Pingpeng Yuan, Yujiang Wang, Jiangji Peng, Tianyu Ma, Siyuan He, Ling Liu

Editor: Multi-Resolution Cleaning of Multivariate Time Series Via Detect-Localize-Repair 3419
Chenyang Li, Chaohong Ma, Xiaohui Yu, Cailong Li, Xiaofeng Meng

GLIDE: GPU-Accelerated ANN Graph Index Construction via Data Locality 3433
Fuhao Ruan, Ziyang Yue, Ling Xu, Dawei Liu, Bolong Zheng

Fast k-Means via Data-Aware Grouping and Gap-Optimized Lower Bound 3447
Xiaogang Huang, Dan Zhuang, Jianbao Chen, Tiefeng Ma, Shuangzhe Liu

BAMG: A Block-Aware Monotonic Graph Index for Disk-Based Approximate Nearest Neighbor Search 3461
Huiling Li, Xin Huang, Byron Choi, Jianliang Xu

Representative Functional Dependencies 3474
Qiongqiong Lin, Jingyan Sai, Jiazheng Song, Jinfei Liu, Kui Ren, Tianzhen Wang, Yanbei Pang, Feifei Li

Unveiling Semantically Cohesive Structures: Maximal Meta-Path Clique Enumeration in Heterogeneous Graphs 3487
Tao Yu, Wen Deng, Weiguo Zheng, Jeffrey Xu Yu

Fast Content-Aware Influence Maximization Query Answering by Labeling Index 3501
Xingliang Lv, Qihao Shi, Can Wang, Mingli Song, Wenliang Du, Wujian Yang, Guanlin Chen

[Experiment, Analysis, and Benchmark] Systematic Evaluation of Plan-Based Adaptive Query Processing 3515
Pei Mu, Anderson Chaves Carniel, Antonio Barbalace, Amir Shaikhha

One Size Does NOT Fit All: on the Importance of Physical Representations for Datalog Evaluation 3529
Nick Rassau, Felix Schuhknecht

Semantic Compression for Sound and Complete Query Answering Over Knowledge Graphs 3542
Junhua Ma, Jianfeng Du, Hai Wan, Yue Yu, Kunxun Qi, Weilin Luo, Yanan Liu

FlashEKGR: Fast Embedding-Based Knowledge Graph Reasoning Models Training 3555
Wentai Zhang, Teng Xu, Weiguang Wang, Junxing Li, Jun Zhang, Yifan Zhu, Haihong E

OMNIA: Closing the Loop by Leveraging LLMs for Knowledge Graphs Completion 3569
Frédéric Ieng, Soror Sahri, Mourad Ouzzani, Massinissa Hammaz, Salima Benbernou Hanieh Khorashadizadeh, Sven Groppe, Farah Benamara

Revisiting Single-Table Retrieval: An Open Problem Under 360° Stress Tests 3583
Chenyu Yang, Ziyu Jiang, Junhao Li, Yuyu Luo, Ju Fan, Nan Tang

Misfeat: Feature Selection for Subgroups With Mutual Information Estimation 3598
Bar Genossar, Thinh On, Md Mouinul Islam, Ben Eliav, Senjuti Basu Roy, Avigdor Gal

Improving Data Imputation Through a Tuned Strategy for Dependency Discovery 3612
Bernardo Breve, Loredana Caruccio, Tullio Pizzuti, Giuseppe Polese

QPAD: Quantile-Preserving Approximate Dimension Reduction for Nearest Neighbors Preservation in High-Dimensional Vector Search 3625
Jiuzhou Fu, Dongfang Zhao

ZTab: Domain-Based Zero-Shot Annotation for Table Columns 3639
Ehsan Hoseinzade, Ke Wang

DLRover-LM: LLM Pre-Training Framework With Thousands of Accelerators in AntGroup 3653
Ziling Huang, Zhengmao Ye, Qingsong Cai, Zelong Huang, Bo Sang, Haitao Zhang, Jian Sha, Tingfeng Lan Hui Lu, Yuanchun Zhou, Mingjie Tang

Tackling Workload Forecasting Challenges with an Offline-Online Dynamic Framework 3668
Jian Jiang, Yu Liu, Jia Li, Lu Han, Wei Lu, Qiwen Deng, Zhibo Zhu, Xingyu Lu, Lintao Ma

OceanBase Mercury: Building a Distributed Real-Time Analytical Processing Database System 3681
Quanqing Xu, Chuanhui Yang, Ruijie Li, Dongdong Xie, Hui Cao, Yi Xiao, Junquan Chen, Saitong Zhao Fusheng Han

High-Fidelity and Complex Test Data Generation for Google SQL Code Generation Services 3694
Shivasankari Kannan, Yeounoh Chung, Amita Gondi, Tristan Swadell, Fatma Özcan

Taxon: Hierarchical Tax Code Prediction with Semantically Aligned LLM Expert Guidance 3708
Jihang Li, Qing Liu, Zulong Chen, Jing Wang, Wei Wang, Chuanfei Xu, Zeyi Wen

TAT: Temporal-Aligned Transformer for Multi-Horizon Peak Demand Forecasting 3722
Zhiyuan Zhao, Sitan Yang, Kin G. Olivares, Boris N. Oreshkin, Stan Vitebsky, Michael W. Mahoney B. Aditya Prakash, Dmitry Efimov

Hierarchical Industrial Demand Forecasting with Temporal and Uncertainty Explanations 3737
Harshavardhan Kamarthi, Shangqing Xu, Xinjie Tong, Xingyu Zhou, James Peters, Joseph Czyzyk B. Aditya Prakash

Automatic Parameter Tuning for Compaction in Lsm-Tree Based Databases 3750
Pinshan Cao, Peng Cai, Xuan Zhou, Jun-Peng Zhu, Kecheng Luo, Sijia Li, Quanqing Xu, Chuanhui Yang

REVISION:Reflective Intent Mining and Online Reasoning Auxiliary for E-Commerce Visual Search System Optimization 3764
Yiwen Tang, Qiuyu Zhao, Zenghui Sun, Jinsong Lan, Xiaoyong Zhu, Bo Zheng

DBdoctor: A Fine-Grained and Non-Intrusive Performance Diagnosis Platform for Databases 3772
Xinyue Shi, Quanqi Xin, Zhengjin Wang, Xinyi Zhang, Haoqiong Bian, Wei Lu, Qiyu Zhuang, Shuang Liu Jikuan Zhang, Xiang Zheng, Yunpeng Chai, Xiaoyong Du

CCD-Level and Load-Aware Thread Orchestration for in-Memory Vector ANNS on Multi-Core CPUs 3785
Yuchen Huang, Baiteng Ma, Yiping Sun, Yang Shi, Xiao Chen, Xiaocheng Zhong, Zhiyong Wang, Yao Hu Chuliang Weng

REG4Rec: Reasoning-Enhanced Generative Model for Large-Scale Recommendation Systems 3799
Haibo Xing, Hao Deng, Yucheng Mao, Lingyu Mu, Jinxin Hu, Yi Xu, Hao Zhang, Jiahao Wang Shizhun Wang, Yu Zhang, Xiaoyi Zeng, Jing Zhang

Graph Query Generation with Constraint-Guided Large Language Agents 3813
Mengying Wang, Nicolaas Jedema, Rahul Pandey, RaviKiran Krishnan, Jens Lehmann, Yinghui Wu

JITPrune: An Efficient Online Feature Pruning Framework for Embedding-Based DLRM Training 3820
Hongzheng Li, Yucheng Wu, Junjie Zhai, Anan Liu, Yuekui Yang, Yingxia Shao

D2SQA: An Edge-Cloud Collaborative Slow Query Analysis Framework Deployed at DBAPPSecurity 3834
Ziquan Fang, Xiangheng Wang, Zijun Jia, Bo Liu, Yuan Fan, Haichuan Zhang

GALA: Generative Aligned Learning for Adaptive Multimodal Representation in the Taobao Shangou Recommender System 3848
Jiping Liu, Zhongmin Zhang, Zisen Sang, Zhijia Fang, Tao Ouyang, Ma Jiang, Shaopeng Liang Zeyang Hou, Guodong Cao, Jia Jia

KScaNN: Scalable Approximate Nearest Neighbor Search on Kunpeng 3861
Oleg Senkevich, Siyang Xu, Tianyi Jiang, Alexander Radionov, Jan Tabaszewski, Dmitriy Malyshev Zijian Li, Daihao Xue, Licheng Yu, Weidi Zeng, Meiling Wang, Xin Yao, Siyu Huang, Gleb Neshchetkin Qiuling Pan, Yaoyao Fu

Efficient Data Processing using On-the-Fly Host-PIM Interactions in a Commodity PIM System 3874
Hyojune Kim, Jeonghyeon Joo, Taehyeong Park, Yongjun Park, Hyuck Han, Sooyong Kang

REDPARROT: Accelerating NL-to-DSL for Business Analytics via Query Semantic Caching 3888
Tong Wang, Yongqin Xu, Jianfeng Zhang, Lingxi Cui, Wenqing Wei, Suzhou Chen, Huan Li, Ke Chen Lidan Shou

CoLIBRi: Supporting Quotation Through Multi-Modal Retrieval and Conversational Search on Manufacturing Drawings 3901
Jacob Pollack, Lucas Peter, Matthias Täschner, Carmen Ahnert

Building and Benchmarking Large Language Models for Machine Translation in Social Network Services 3915
Hongcheng Guo, Fei Zhao, Shaosheng Cao, Xinze Lyu, Zijie Meng, Yue Wang, Yao Hu, Zhoujun Li Zuozhu Liu

FELA: A Multi-Agent Evolutionary System for Feature Engineering of Industrial Event Log Data 3924
Kun Ouyang, Haoyu Wang, Dong Fang

Cascading Relevance-Driven Recommendation Network for CTR Prediction in Trigger-Introduced Recommendation 3938
Kaixuan Chen, Wenwen Wang, Xing Fang, Yang Huang, Jing Wang

Accurate and Efficient Multi-Channel Time Series Forecasting via Sparse Attention Mechanism 3945
Lei Gao, Hengda Bao, Jingfei Fang, Guangzheng Wu, Weihua Zhou, Yun Zhou

From Benchmarks to Production: Transferring Time Series Anomaly Detection Methods for Electricity Production Monitoring 3958
Nicolas Vautier, Paul Caron, Nardi Xhepi, Félicie Bizeul, Manel Boumghar, Christophe Degouy Paul Boniol

OpenZL: Using Graphs to Compress Smaller and Faster 3971
Yann Collet, Nick Terrell, Winston Felix Handte, Danielle Rozenblit, Victor Zhang, Kevin Zhang Yaelle Goldschlag, Jennifer Lee, Elliot Gorokhovsky, Yonatan Komornik, Daniel Riegel, Stan Angelov Nadav Rotem

GaV: Guess and Verification of Column Semantics 3985
Davide Di Stefano, Jinsong Guo, Yang Hu, Matteo Capalbo, Davide Mario Longo, Georg Gottlob

User-Adaptive Meta-Learning for Cold-Start Medication Recommendation with Uncertainty Filtering 3992
Arya Hadizadeh Moghaddam, Mohsen Nayebi Kerdabadi, Dongjie Wang, Mei Liu, Zijun Yao

On Efficient Materialization in Data Lakes 4006
Andrew Harn, Herald Kllapi, Zhepeng Yan

Billion-Scale Fintech Analytics: Scalable Data Management and Anomaly Detection at NPCI 4020
Bharadwaj Dasari, Turaga Sai Dhiraj, Ganesh Jambhrunkar, Thirumalai Kailasam, Charu Vikram Saurav Singla, Pranjal Naman, Yogesh Simmhan

Decoupled Multimodal Fusion for User Interest Modeling in Click-Through Rate Prediction 4034
Alin Fan, Hanqing Li, Sihan Lu, Jingsong Yuan, Jiandong Zhang

GalaxyRAG: Graph Retrieval-Augmented Generation for Enterprise Knowledge Systems 4047
Bing Tong, Yan Zhou, Chen Zhang, Zhaojie Yin, Jia Li

OEPO: Online Experience-based Preference Optimization for CTR Prediction 4060
Zhichao Liao, Ziheng Ni, Congcong Liu, Zhiwei Fang, Changping Peng

DM-RAG: Enhancing User Support in Dameng Databases with Retrieval-Augmented Generation 4067
Qiang Huang, Ke Liu, Liang Deng, Sijing Zhang, Chuang Hu, Tieyun Qian, Xiao Yan, Jiawei Jiang

Relevance Matters: A Multi-Task and Multi-Stage Large Language Model Approach for E-Commerce Query Rewriting 4081
Aijun Dai, Jixiang Zhang, Haiqing Hu, Guoyu Tang, Lin Liu, Ziguang Cheng

Bala-Join: An Adaptive Hash Join for Balancing Communication and Computation in Geo-Distributed SQL Databases 4089
Wenlong Song, Hui Li, Bingying Zhai, Jinxin Yang, Pinghui Wang, Luming Sun, Ming Li, Jiangtao Cui

The Evolution of LSM-Tree Key-Value Stores: A Tutorial on State-Of-The-Art and Future Directions 4103
Yina Lv, Qiao Li, Quanqing Xu, Chun Jason Xue

Large Language Models for Spatial Analysis Queries 4108
Mohamed Hemdan, Youssef Hussein, Mohamed F. Mokbel

Data Discovery in Data Lakes: Operations, Indexes, Systems 4114
Ziawasch Abedjan, Mahdi Esmailoghli, Sainyam Galhotra

The Virtuous Cycle: AI-Powered Vector Search and Vector Search-Augmented AI 4120
Jiuqi Wei, Quanqing Xu, Chuanhui Yang

Query Rewrite in the Learning Age: From Rules to ML-Based and LLM-Driven Techniques 4127
Shengchen Liu, Verena Kantere, Nicholas Ostan, Farhana Haider, Calisto Zuzarte

Data-Centric Foundations of Agentic AI 4132
Yuxin Jin, Hanchen Wang, Ying Zhang, Wenjie Zhang

Model Slicing: A Data Engineering Perspective 4138
Parke Godfrey, Lukasz Golab, Divesh Srivastava, Jarek Szlichta

BClean+: A Bayesian Data Cleaning System with Automated Prior Generation 4143
Ziyan Han, Jing Zhu, Jinbin Huang, Rui Mao, Jianbin Qin

LazyVLM: Neuro-Symbolic Approach to Video Analytics 4147
Xiangru Jian, Wei Pang, Zhengyuan Dong, Chao Zhang, M. Tamer Özsu

Jazero: A Semantic Table Search System 4151
Martin Pekár Christensen, Matteo Lissandrini, Katja Hose

GeX: Guiding Tuning of Db2 With Explainable AI 4155
Andrew Chai, Alexander Bianchi, Vincent Corvinelli, Parke Godfrey, Lukasz Golab, Jarek Szlichta Calisto Zuzarte

A Fast, Versatile, and User-Friendly Plugin for Kernel Density Analysis 4160
Tsz Nam Chan, Bojian Zhu, Leong Hou U, Dingming Wu, Wei Tu, Jianliang Xu

Schema-GraphRAG: Bridging Hybrid Search and Graph Traversal for Complex Retrieval Tasks 4164
Bastian Lipka, Venkata Vamsikrishna Meduri, Berthold Reinwald, Nasrullah Sheikh

TAPE: A Temporal Graph-Based Memory System for Personal LLM Agents 4168
Chengyang Luo, Qing Liu, Wenjie Zhang, Yunjun Gao

Scoper: Streamline Linkable Schemas for Matching 4172
Leonard Traeger, Andreas Behrend, George Karabatis

CORAL: COncept-Based Explanations for RAG LLMS 4176
Katherine Ling, Joel Rorseth, Parke Godfrey, Lukasz Golab, Jarek Szlichta

DeepSketch 2.0: Discovering Temporal Relationships in Large Time Series Datasets 4181
Zheng Zhang, Runzhe Jiang, Andrew Crotty

DJGen: Data & Conjunctive Join Plan Generator 4185
Vasilis Sarris, Brian T. Nixon, Panos K. Chrysanthis

Case: A Comprehensive and Interactive Influence Analysis System for Social Networks 4189
Xueqin Chang, Chuanyu Liu, Qing Liu, Baihua Zheng, Yunjun Gao

HEX: Olap-Enabled Hierarchical Explanations 4193
Kathryn Carbone, Parke Godfrey, Lukasz Golab, Jarek Szlichta, Robin Cohen

RUBEN: Rule-Based Explanations for Retrieval-Augmented LLM Systems 4198
Joel Rorseth, Parke Godfrey, Lukasz Golab, Divesh Srivastava, Jarek Szlichta

SENTINEL: Evaluating Pipeline Robustness to Distributional Shifts 4203
Jahid Hasan, Jingya Wang, Romila Pradhan

SqlRewriter: Harnessing Community Knowledge to Rewrite SQL Queries 4207
Qiushi Bai, Yihong Yu, Colin Michael Harrison, Jiatong Liu, James Liu, Jessie He, Hartley Tran, Jun Xia Chen Li

LOADs: Adaptive Cloud-Edge-Device Database Management System Optimizer 4211
Chunyu Zhao, Yihan Zhang, Shuangshuang Cui, Hongzhi Wang

Multi-Model Geospatial Data Management and Exploration 4215
David Lengweiler, Marco Vogt, Heiko Schuldt

PIPER - Leveraging Pipeline Perspectives for Effective Data Pipeline Exploration 4219
Ridhwan Hakim Bin Kusni, Melanie Herschel

Pathfinder: Context Engineering and Knowledge Management for Domain-Specific Horizontal Reasoning 4223
Joohyun Lee, Ghita Benboubker, JungKwan Han, Jisoo Jang, Wen-Syan Li

Efficient Neural-Symbolic Data System via Multi-Agent Collaboration 4227
Ye Yuan, Bo Tang, Zhaojing Luo, Boyang Li, Renjie Liu, Zhilang Wei

VPS: Rethinking OLTP Database Performance Evaluation Through Transactional Value 4239
Jianbin Qin, Yibin Lin, Wendi Hua, Rui Mao, Chuan Xiao

Living Databases: A Unified Model for Continuous Schema Evolution, Versioning, and Transformations 4247
Amol Deshpande

Green or Greedy? An Ecological Analysis of Data Center GPU Replacements 4256
Marc Baeuerle, Ole Becker, Nikolas Hoellerl, Ricardo Salazar Díaz, Ilin Tolovski, Tilmann Rabl

GenIE: Simulator-Driven Iterative Data Exploration for Scientific Discovery 4266
Ashwin Gerard Colaco, Martin Boissier, Sriram Rao, Shubharoop Ghosh, Sharad Mehrotra, Tilmann Rabl

Towards a Hybrid Quantum-Classical Computing Framework for Database Optimization Problems in Real Time Setup 4278
Hanwen Liu, Ibrahim Sabek

Minimum k-Vertex Connected Graph Search (Extended Abstract) 4289
Yang Liu, Hejiao Huang, Kaiqiang Yu, Shengxin Liu, Cheng Long

Orthogonal Keys: High Precision and Recall for Mining Meaningful Database Keys from Inconsistent and Incomplete Relations (Extended Abstract) 4291
Henning Köhler, Sebastian Link

Hierarchy-Aware Neural Subgraph Matching with Enhanced Similarity Measure (Extended Abstract) 4293
Zhouyang Liu, Ning Liu, Yixin Chen, Jiezhong He, Menghan Jia, Dongsheng Li

Graph2Region: Efficient Graph Similarity Learning With Structure and Scale Restoration (Extended Abstract) 4295
Zhouyang Liu, Yixin Chen, Ning Liu, Jiezhong He, Dongsheng Li

Maximizing Influence Query Over Indoor Trajectories (Extended Abstract) 4297
Jian Chen, Hong Gao, Yuhong Shi, Junle Chen, Donghua Yang, Jianzhong Li

An Amortized O(1) Lower Bound for Dynamic Time Warping in Motif Discovery (Extended Abstract) 4299
Zemin Chao, Hong Gao, Dongjing Miao, Jianzhong Li, Hongzhi Wang

JetBGC: Joint Robust Embedding and Structural Fusion Bipartite Graph Clustering (Extended Abstract) 4301
Liang Li, Yuangang Pan, Junpu Zhang, Pei Zhang, Jie Liu, Xinwang Liu, Kenli Li, Ivor W. Tsang, Keqin Li

ConvD: Attention Enhanced Dynamic Convolutional Embeddings for Knowledge Graph Completion (Extended Abstract) 4303
Wenbin Guo, Zhao Li, Xin Wang, Zirui Chen, Jun Zhao, Jianxin Li, Ye Yuan

Snoopy: Effective and Efficient Semantic Join Discovery Via Proxy Columns (Extended Abstract) 4305
Yuxiang Guo, Yuren Mao, Zhonghao Hu, Lu Chen, Yunjun Gao

Discovery of Temporal Network Motifs (Extended Abstract) 4307
Hanqing Chen, Shuai Ma, Junfeng Liu, Lizhen Cui

LIOF: Make the Learned Index Learn Faster with Higher Accuracy (Extended Abstract) 4309
Tao Ji, Kai Zhong, Luming Sun, Yiyan Li, Cuiping Li, Hong Chen

FedDict: Towards Practical Federated Dictionary-Based Time Series Classification (Extended Abstract) 4311
Zhiyu Liang, Zheng Liang, Hongzhi Wang, Bo Zheng

Generalized Local Prominence for Source Detection in Real-World Rumor Networks (Extended Abstract) 4314
Syed Shafat Ali, Ajay Rastogi, Tarique Anwar, Syed Afzal M. Rizvi, Jian Yang, Jia Wu, Quan Z. Sheng

Intent Propagation Contrastive Collaborative Filtering Extended Abstract 4316
Haojie Li, Junwei Du, Guanfeng Liu, Feng Jiang, Yan Wang, Xiaofang Zhou

HyCubE: Efficient Knowledge Hypergraph 3D Circular Convolutional Embedding (Extended Abstract) 4318
Zhao Li, Xin Wang, Jun Zhao, Wenbin Guo, Jianxin Li

MDSD: Multi-turn Diverse Synthetic Dialog Generation for Domain Specific Incomplete Requests Understanding 4320
Xi Li, Xiaoxu Wu, Lijuan Xiao, Tao Liu, Ping Huang, Jiulong Shan Apple

Model Slicing: A Data Engineering Perspective 4322
Parke Godfrey, Lukasz Golab, Divesh Srivastava, Jarek Szlichta

Responsible Entity Resolution Over Streaming Data 4324
Kostas Stefanidis, Vasilis Efthymiou, Tiago Brasileiro Araújo

Tuning IBM Db2 with Explainable AI 4326
Andrew Chai, Alexander Bianchi, Vincent Corvinelli, Parke Godfrey, Lukasz Golab, Jarek Szlichta Calisto Zuzarte

From Polystores to Polydbms: The Polypheny Experience 4328
Marco Vogt, David Lengweiler, Martin Vahlensieck, Yiming Wu, Heiko Schuldt

Recovering Structure in Unstructured LLM Outputs 4329
Joel Rorseth, Parke Godfrey, Lukasz Golab, Divesh Srivastava, Jarek Szlichta

Rethinking BFT Consensus via Transaction-Level Protocol Construction for Optimal Performance 4331
Bingxu Zhu, Gengrui Zhang

Evaluating Pipeline Robustness Under Distributional Shifts 4333
Jahid Hasan, Jingya Wang, Romila Pradhan

SPLICE: An Efficient Framework for Selecting Sources for Machine Learning Tasks 4334
Ambarish Singh, Romila Pradhan

When Text-to-SQL Evaluation Misleads: Rethinking Benchmarking Practices 4335
Oktie Hassanzadeh, Yotam Perlitz, Nhan Pham, Timothy Dinger, Tanvi Kaple, Long Vu, Michael Glass Dharmashankar Subramanian

Is Quantum Computing Ready for Real-Time Database Optimization? 4336
Hanwen Liu, Ibrahim Sabek

On Breaking the Scalability Barrier in Data Cleaning 4338
El Kindi Rezig

GNN Explainers 2.0: A Paradigm for User-Oriented, Data-Guided Explanations 4340
Arijit Khan

SalesforceDB: Built on One LSM to Rule Them All ! 4342
Vaibhav Arora, Subho Chatterjee, Nat Wyatt