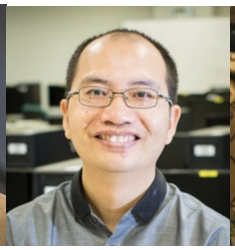
Table of Contents

A MESSAGE FROM THE CHAIRS

GENERAL CO-CHAIRS PROGRAM COMMITTEE CO-CHAIRS

Malu Castellanos
Teradata, USA

Yannis Velegrakis
Utretch University, The Netherlands



Bingshen He
National Univ. of of Singapore, Singapore

Gautam Das
Univ. of Texas at Arlington, USA

Timos Sellis
Athena Research Center, Greece

ICDE is IEEE's TCDE flagship conference, focused on the efficient management of data. Now in its 40th edition, ICDE continues being a premier forum for data management and database researchers, vendors, practitioners, application developers, and users. This year it takes place in the beautiful city of Utretch, The Netherlands. Data management in The Netherlands is an active topic of research and there are 12 research groups covering topics like Data Systems, Quality and Integration, Knowledge Representation, Provenance, Data Mining, Exploration, Information Retrieval and Data Analytics. A special Symposium (DBDBD) is organized yearly where Data Management Researchers from both Netherlands and Belgium meet and present their recent works. The Dutch government has been constantly investing millions of Euros yearly in AI and Data Science and this trend is expected to remain.

The data management field is witnessing rapid development in many areas, and ICDE offers a forum to learn, share, and discuss these advances. To capture this reality, this year's ICDE has an exciting program that includes three keynotes covering a wide range of interesting topics ranging from "*Duality: Converging the Worlds of Relational, Documents, and Graph to Simplify App Dev*" by Juan Loaiza (EVP Oracle) and Tirthankar Lahiri (SVP Oracle), to *"How serious are we about green computing? The impact of data intensive computing" by* Gustavo Alonso (Professor, ETH), to *"AI Systems beyond Accelerating Linear Algebra"* by Christos Kozyrakis (Professor, Stanford University). A panel on *"Data Management in the Cloud: Trends and Directions"* moderated by C. Mohan with renowned researchers from the data management area. A special track on Data Engineering Future Technologies (DEFT) chaired by Chair Beng Chin Ooi with 13 forward looking papers that provide visions for data engineering and analytics in the near future. And a Lightning Talks session (introduced in ICDE 2018) with 15 talks that present early-stage innovative ideas, chaired by Minos Garofalakis.

In keeping with the tradition, the conference also includes a research track consisting of the presentations of 376 papers, an industrial and application track with 21 papers, a demo track with 27 demonstrations, and a selected set of 7 tutorials, 8 workshops, and a PhD symposium..

The ICDE 2024 conference would not have been possible without the efforts of a large number of members of our community participating in the Technical Program Committee and the Organization Committee.

We would like to extend our sincere appreciation to the keynote speakers Juan Loaiza, Tirthankar Lahiri, Gustavo Alonso and Christos Kozyrakis, and to our panel moderator C. Mohan. Our deepest gratitude to the following Chairs for their outstanding and conscientious work in helping us put together an excellent program: Research Program Chairs BingSheng He, Gautam Das and Timos Sellis; Future Technologies Chair Beng Chin Ooi; Industry and Application Chairs Ashraf Aboulnaga, Fatma Ozcan and Peter Boncz; Demonstration Chairs Angela Bonifati, Asterios Katsifodimos, Wookey Lee; Lightning Talks Chair Minos Garofalakis; Tutorial Chairs Felix Naumann, Renata Borovica-Gajic and Xiaofang Zhou; Workshop Chairs George Fletcher and Steven Whang; PhD Symposium Chairs Alekh Jindal and Byron Choi; TKDE Posters Chair Shuhao Zhang. We also would like to extend our sincere gratitude for the excellent work to the rest of the Organization Committee: Proceedings Chair Odysseas Papapetrou; Sponsorship Chairs Yeye He, Torben Bach Pedersen and Zhiyong Peng; Registration Chair Rihan Hai; and Diversity and Inclusion Chairs Katerina Ioannou and Pinar Tozun.

The Program Committee chairs would especially like to thank the Area Chairs and members of the Technical Program Committee who, in spite of the unprecedented number of submissions (almost twice the number of submissions compared to ICDE 2023) and tight timelines, did an outstanding job in the reviews and discussions, thus ensuring that ICDE continues to live up to its reputation of being a premier database conference. This year out of the 376 papers accepted in the research track,10 were accepted directly, 357 were accepted after revision, and 9 were accepted after a shepherding phase.

We want to thank all authors of submitted papers for their interest in the conference and to the authors of accepted papers for traveling to Utretch to share their work with the ICDE community.

Special thanks to the local volunteers of the Data Intensive Systems group and the administrative personnel of Utrecht University, as well as those of the neighboring universities, the NSF grant recipients, and everyone else that helped in making this conference a success.

Finally, we are grateful to the financial support from the U.S. National Science Foundation and TCDE's contribution to enable student financial grants, and to all our sponsors whose support was essential to make this conference possible. We also want to thank the CMT team for their support with the submission site.

We welcome you to ICDE 2024 in Utrecht and hope you enjoy the conference!

General Chairs

Malu Castellanos, Teradata, USA
Yannis Velegrakis, Utrecht University, Netherlands

PC Chairs

BingSheng He, NUS, Singapore
Gautam Das, University of Texas at Arlington, USA
Timos Sellis, Athena Research Center, Greece

Conference Officers

General Chairs
Malu Castellanos, Teradata, USA
Yannis Velegrakis, Utrecht University, Netherlands

PC Chairs
Gautam Das, University of Texas at Arlington, USA
Timos Sellis, Athena Research Center, Greece
BingSheng He, NUS, Singapore

Industry and Applications Chairs
Fatma Ozcan, Google, USA
Peter Boncz, CWI, Netherlands
Ashraf Aboulnaga, QCRI, Qatar

Demonstrations Chairs
Asterios Katsifodimos, TU Delft, Netherlands
Angela Bonifati, Lyon 1 University, France
Wookey Lee, INHA University, South Korea

Tutorial Chairs
Felix Naumann, HPI, Germany
Renata Borovica-Gajic, University of Melbourne, Australia
Xiaofang Zhou, HKUST, China

Special Session Data Engineering Future Technologies Chair
Beng Chin Ooi, NUS, Singapore

Lightning Talks Chair
Minos Garofalakis, Athena Research Center, Greece

Workshop Chairs
George Fletcher, TU Eindhoven, Netherlands
Steven Whang, KAIST, South Korea

Panel Chair
C. Mohan, Hong Kong Baptist University, Hong Kong

PhD Symposium Chair
Alekh Jindal, SmartApps, USA
Byron Choi, Hong Kong Baptist University, Hong Kong

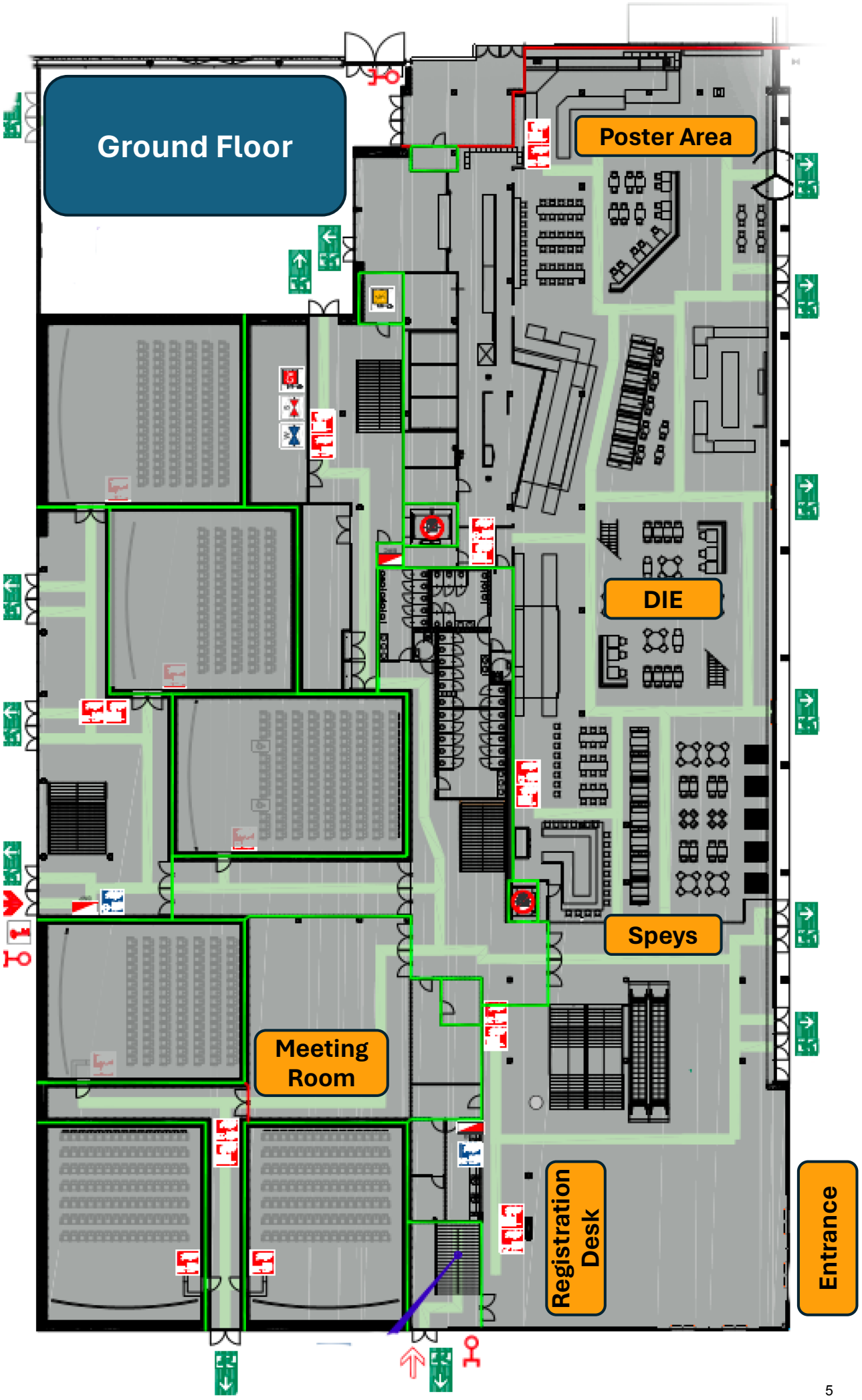
Publication and Proceedings Chair
Odysseas Papapetrou, TU Eindhoven, Netherlands

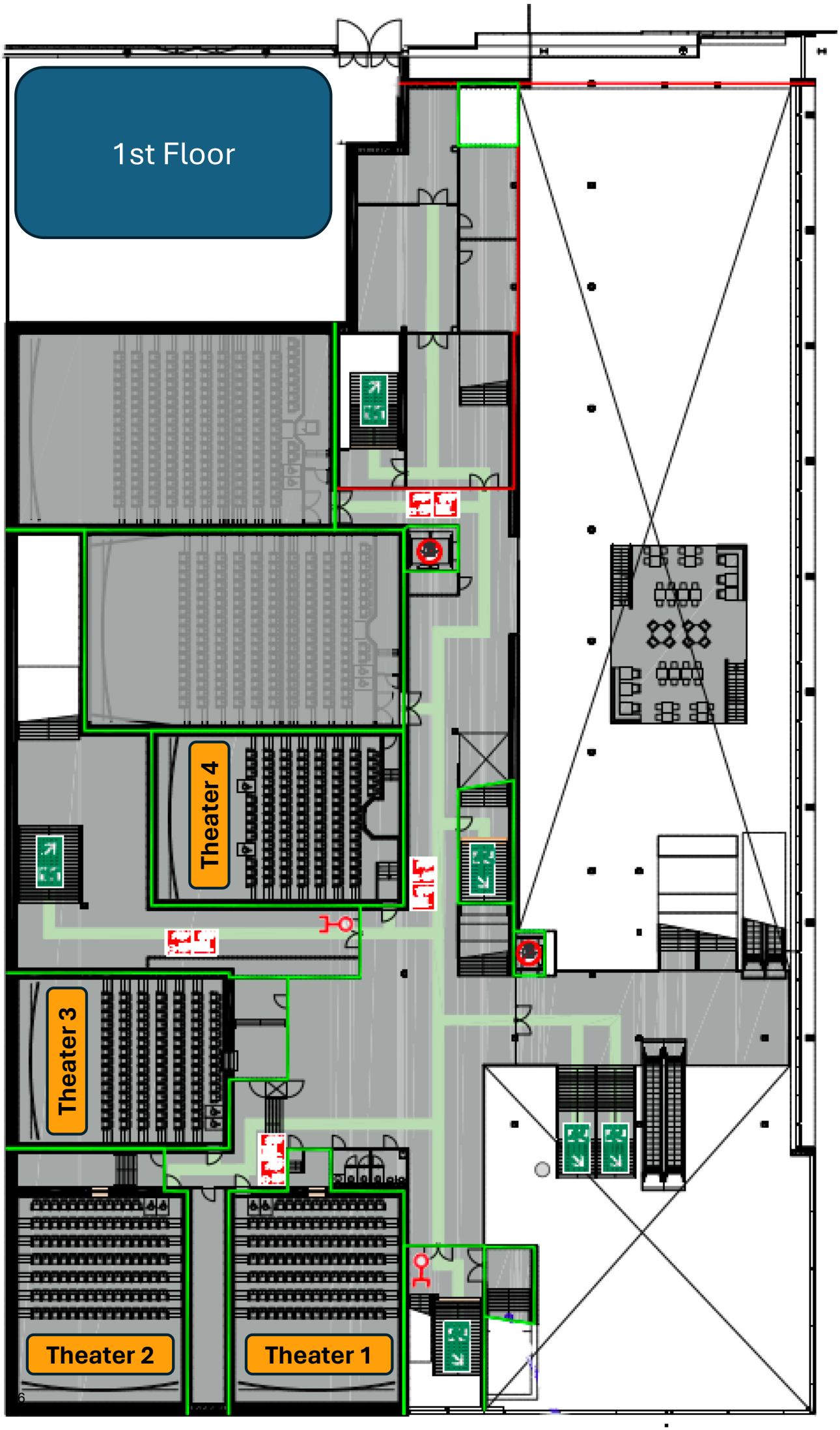
Sponsorship Chairs
Yeye He, Microsoft, USA
Torben Bach Pedersen, Aalborg University, Denmark
Zhiyong Peng, CCF TCDB, China

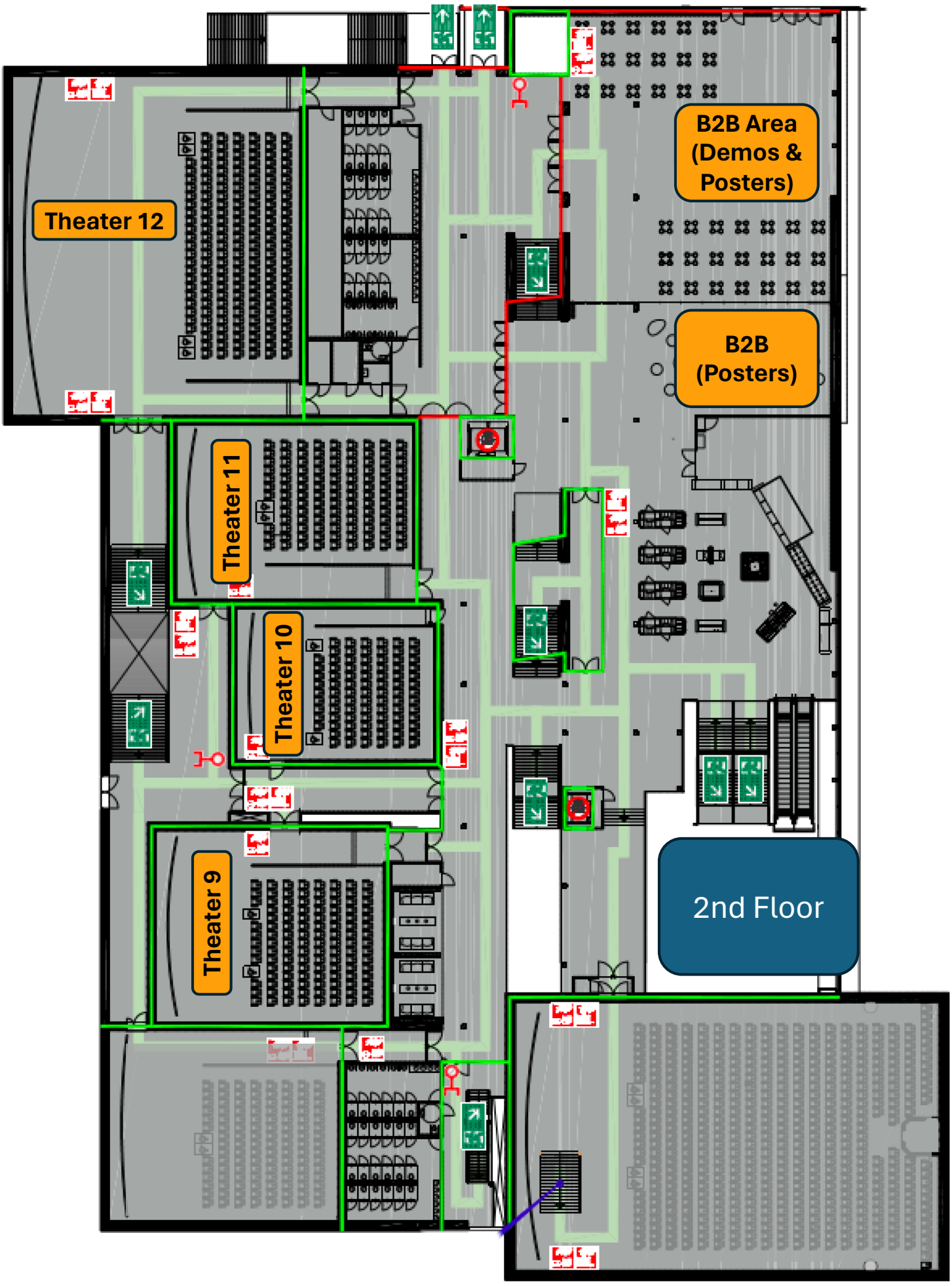
Diversity and Inclusion Chairs
Katerina Ioannou, Tilburg University, Netherlands
Pinar Tozun, IT University of Copenhagen, Denmark

Registration Chair
Rihan Hai, TU/Delft, Netherlands

TKDE Posters Chairs
Shuhao Zhang, Nanyang Technological University


Ground Floor
Poster Area
DIE
Speys
Meeting Room
Registration Desk
Entrance


1st Floor
Theater 4
Theater 3
Theater 2
Theater 1


Theater 12
B2B Area (Demos & Posters)
B2B (Posters)
Theater 11
Theater 10
Theater 9
2nd Floor

Monday May 13th, 2024

	Theater 12	Theater 11	Theater 10	Theater 9	Theater 4	Theater 3	Theater 2	Theater 1	Speys	B2B
9:00-10:00		DataPlat	HarDB & Active	DECOR	FAIR	MulTiSA	DBML	SEAGraph		
10:00-10:30	Coffee Break @ Speys									
10:30-12:00		DataPlat	HarDB & Active	DECOR	FAIR	MulTiSA	DBML	SEAGraph		
12:00-13:30	Lunch @ Speys									
13:30-15:00		DASC	HarDB & Active	DECOR	FAIR	MulTiSA	DBML	SEAGraph		
15:00-15:30	Coffee Break @ Speys									
15:30-17:00		DASC	HarDB & Active	DECOR	FAIR	MulTiSA	DBML	SEAGraph		
17:00-19:30	Welcome Reception @ Speys									

Tuesday May 14th, 2024

	Theater 12	Theater 11	Theater 10	Theater 9	Theater 4	Theater 3	Theater 2	Theater 1	Speys	B2B
8:30-8:45	Opening @ Theater 12									
8:45-9:45	Keynote on Duality: Converging the Worlds of Relational, Documents, and Graph to Simplify App Dev by Juan Loaiza, (Executive Vice President, Oracle) and (Tirthankar Lahiri is Senior Vice President of Data, Oracle). @ Theater 12									
9:45-10:15	Coffee Break @ Speys									
10:15-11:45	AI for Databases I	Query Processing, Indexing, and Optimization I	Graphs, Networks, and Semistructured Data I	Data Mining and Knowledge Discovery I	Very Large Data Science Applications/pipelines	Information Integration and Data Quality I	Cloud Database Systems [Industry]. Includes invited talk on Vector Search and Databases	Tutorial on Bipartite Graph Analytics: Current Techniques and Future Trends	Posters of the 15:45 Session	Posters of the 15:45 Session
11:45-13:15	Lunch @ Speys									
13:15-15:15	Sponsor Talks @ Theater 12									
15:15-15:45	Coffee Break @ Speys									
15:45-16:21	AI for Databases II	Query Processing, Indexing, and Optimization II	Graphs, Networks, and Semistructured Data II	Data Mining and Knowledge Discovery II	Spatial Databases and Temporal Databases I	Information Integration and Data Quality II	Data Mining and Knowledge Discovery IV	Data Mining and Knowledge Discovery V	Posters of the 10:30 Session	Demonstration Group A and B and PhD Posters
16:21-18:00	AI for Databases III	Query Processing, Indexing, and Optimization III	Graphs, Networks, and Semistructured Data III	Data Mining and Knowledge Discovery III	Spatial Databases and Temporal Databases II	Information Integration and Data Quality III	Machine Learning and Data Science [Industry]	Tutorial on LLMs: Principles and Practice	Posters of the 10:30 Session	Demonstration Group A and B and PhD Posters
18:00-20:00	TCDE Reception @ Speys								TKDE Posters	TKDE Posters

Wednesday May 15th, 2024

	Theater 12	Theater 11	Theater 10	Theater 9	Theater 4	Theater 3	Theater 2	Theater 1	Speys	B2B
8:30-9:30	Keynote on How Serious We Are About Green Computing: The Impact of Data Intensive Computing by Gustavo Alonso (Professor, ETH Zurich) @ Theater 12									
9:30-10:00	Coffee Break @ Speys									
10:00-12:15	Database technology for AI I	Distributed, Parallel and P2P Data Management I	Explainability, Fairness, and Trust in Data Systems and Analysis I	Data Mining and Knowledge Discovery VI	Crowdsourcing I	Diversity, Equality and Inclusion in Database Venues	Infrastructure for Machine Learning [Industry]. Includes invited talk on Deletion Vectors: No-Regrets Row-Level Updates in Delta Lake	Tutorial on Privacy-Aware Analysis based on Data Series	Posters of the 15:30 Session	Demonstration Group B and C and PhD Posters
12:15-13:45	Lunch @ Speys									
13:45-15:15	Executive and Awards Session @ Theater 12									
15:30-16:30	Database technology for AI II	Distributed, Parallel and P2P Data Management II	Explainability, Fairness, and Trust in Data Systems and Analysis II	Data Mining and Knowledge Discovery VII	Crowdsourcing II	Database technology for Blockchains I	Data Mining and Knowledge Discovery VIII	Data Mining and Knowledge Discovery IX	Posters of the 17:00 Session	Posters of the 17:00 Session
16:30-17:00	Coffee Break @ Speys									
17:00-18:30	Text, Semi-Structured Data, IR, Image, and Multimedia databases I	Graphs, Networks, and Semistructured Data IV	Explainability, Fairness, and Trust in Data Systems and Analysis III	Data Mining and Knowledge Discovery X	Crowdsourcing III	Panel on Data Management in the Cloud: Trends and Directions	Query Performance [Industry]. Includes invited talk on How I Learned to Stop Worrying About Benchmarks	Tutorial on Privacy-Aware Analysis based on Data Series (Part B)	Posters of the 10:00 Session	Posters of the 10:00 Session
19:00-22:30	Banquet Dinner @ Spoorwegmuseum (Railway Museum) or Boat Tour									

Thursday May 16th, 2024

	Theater 12	Theater 11	Theater 10	Theater 9	Theater 4	Theater 3	Theater 2	Theater 1	Speys	B2B
8:30-9:30	Keynote by Christos Kozyrakis (Stanford University) @ Theater 12									
9:30-10:00	Coffee Break @ Speys									
10:00-11:50	Future Technologies 1 @ Theater 12									
11:50-13:15	Lunch @ Speys									
13:15-14:45	Database technology for Blockchains II	Modern Hardware and In-Memory Database Systems I	Graphs, Networks, and Semistructured Data V	Graphs, Networks, and Semistructured Data VI	PhD Symposium	Lightning Talks	Tuning and Industrial Applications [Industry]	Tutorial on Robust Query Optimization in the Era of Machine Learning: State-of-the-Art and Future Directions	Posters of the 16:30 Session	Posters of the 16:30 Session
15:00-16:00	Graphs, Networks, and Semistructured Data VII	Modern Hardware and In-Memory Database Systems II	Graphs, Networks, and Semistructured Data VIII	Graphs, Networks, and Semistructured Data IX	Data Stream Systems and Edge Computing I	Data Mining and Knowledge Discovery XI	Database Security and Privacy I	Data Mining and Knowledge Discovery XIII	Posters of the 10:00 Session	Demonstration Group A and C
16:00-16:30	Coffee Break @ Speys									
16:30-18:40	Graphs, Networks, and Semistructured Data X	Modern Hardware and In-Memory Database Systems III	Graphs, Networks, and Semistructured Data XI	Graphs, Networks, and Semistructured Data XII	Data Stream Systems and Edge Computing II	Data Mining and Knowledge Discovery XII	Database Security and Privacy II	Tutorial on Quantum Data Management: From theory to Opportunities	Posters of the 13:15 and 15:00 Session	Posters of the 13:15 and 15:00 Session

Friday May 17th, 2024

	Theater 12	Theater 11	Theater 10	Theater 9	Theater 4	Theater 3	Theater 2	Theater 1	Speys	B2B
8:30-10:00	Text, Semi-Structured Data, IR, Image, and Multimedia databases II	Query Processing, Indexing, and Optimization IV	Graphs, Networks, and Semistructured Data XIII	Uncertain, Probabilistic, and Approximate Databases	Benchmarking, Performance Modeling, Tuning, and Testing	Workflows and Scientific Data Management	Data Mining and Knowledge Discovery XIV	An Interactive Dive (Tutorial) into Time-Series Anomaly Detection	Posters of the 16:20 Session	Posters of the 16:20 Session
10:00-10:30	Coffee Break @ Speys									
10:30-12:45	Future Technologies 2 @ Theater 12									
12:45-14:00	Lunch @ Speys									
14:00-15:50	AI for Databases IV	Query Processing, Indexing, and Optimization V	Graphs, Networks, and Semistructured Data XIV	Data Mining and Knowledge Discovery XV	Spatial Databases and Temporal Databases III	Information Integration and Data Quality IV	Data Mining and Knowledge Discovery XVII	A Comprehensive Tutorial on the over 100 years of Diagrammatic Representations of Logical Statements and Relational Queries	Posters of the 8:30 and 10:30 Sessions	Posters of the 8:30 and 10:30 Sessions
15:50-16:20	Coffee Break @ Speys									
16:20-18:30	AI for Databases V	Query Processing, Indexing, and Optimization VI	Graphs, Networks, and Semistructured Data XV	Data Mining and Knowledge Discovery XVI	Spatial Databases and Temporal Databases IV	Information Integration and Data Quality V	Data Mining and Knowledge Discovery XVIII	A Comprehensive Tutorial on the over 100 years of Diagrammatic Representations of Logical Statements and Relational Queries (Part B)	Posters of the 14:00 Session	Posters of the 14:00 Session

Co-located Workshops

DECOR'24

7th Int. Workshop on Data Engineering Meets Intelligent Food and Cooking Recipes (DECOR'24)

The DECOR workshop aims to provide a leading forum for researchers, practitioners, developers, and users to explore cutting-edge ideas and to exchange techniques, tools, and experiences in the intersection of Data Engineering and Intelligent Food and Cooking Recipes.

DASC'24

International Workshop on Data-Driven Smart Cities (DASC '24)

The DASC workshop aims to gather professionals from municipalities, industry, and academia focusing on different aspects of Smart city. The purpose of the session is to discuss the latest scientific results and practical use cases, identify the opportunities and challenges of novel best methods and practices on data-driven Smart cities.

FAIR'24

International Workshop on Fairness in AI (FAIR'24)

The Fairness in AI workshop aims at discussing a set of new Machine Learning approaches, applications and paradigms which enhance fairness and explainability in AI algorithms and Systems. It also aims at raising the awareness of the Data Engineering community on the challenges and opportunities of the fairness in AI arena.

MulTiSA'24

International Workshop on Multivariate Time Series Analytics (MulTiSA'24)

The MulTiSA workshop will bring together researchers and practitioners working with multivariate time series, for presenting and discussing requirements, open problems, and related work, and to foster collaborations and further developments in the topic.

DataPlat'24

International Workshop on Data Platform Design, Management, and Optimization (DataPlat'24)

The DataPlat workshop aims at funneling efforts toward the development of data platforms as data-intensive ecosystems supporting data scientists and architects at a high level, and fosters innovative research solutions that contribute to further advancements in this field.

DBML'24

International Workshop on Databases and Machine Learning (DBML'24)

The DBML workshop aims at bringing together researchers and practitioners in the intersection of DB and ML research, providing a forum for DB-inspired or ML-inspired approaches addressing challenges encountered in each of the two areas.

SEAGraph'24

3rd Workshop on Search, Exploration, and Analysis in Heterogenous Datastore, Graph Data Edition (SEAGraph'24)

The SEAGraph workshop proposes a unique international venue for researchers and practitioners willing to share their insights, experience, and solutions in the management and analysis of heterogeneous Graph data. The SEA Graph workshop will provide a forum for researchers and practitioners to exchange ideas, results, and visions on challenges in adopting graphs to handle data management, information extraction, exploration, and analysis of heterogeneous data and multiple data models at once.

HardBD & Active'24

Joint International Workshop on Big Data Management on Emerging Hardware and Data Management on Virtualized Active Systems (HardBD & Active)

The aim of the HardBD&Active workshop is to bring together researchers, practitioners, system administrators, and others who are interested to share their perspectives on exploiting new hardware technologies for data-intensive workloads and big data systems, and to discuss and identify future directions and challenges in this area.

Monday May 13th, 2024 17:00-19:30

Welcome Reception - [In Speys]
The ICDE organizing committee is inviting you for a welcome drink at the Speys restaurant. Please join us for some relaxing time, meet colleagues and friends, and catch up on your their recent works (and not only).

Tuesday May 14th, 2024 8:30-8:45

Opening - [In Theater 12]

Tuesday May 14th, 2024 8:45-9:45

Keynote 1 [In Theater 12, Chair: Malu Castellanos]

Duality: Converging the Worlds of Relational, Documents, and Graph to Simplify App Dev
by Juan Loaiza, (Executive Vice President, Oracle) and Tirthankar Lahiri (Senior Vice President of Data, Oracle).

Juan Loaiza is Executive Vice President of mission-critical database technologies at Oracle. He is responsible for leading product strategy and engineering for the world's leading transaction processing and engineered systems technologies. Juan holds BS and MS degrees in computer science from the Massachusetts Institute of Technology.

Tirthankar Lahiri is Senior Vice President of Data, In-Memory and AI Vector technologies at Oracle. He is responsible for the Oracle Database Data Engine as well as the TimesTen In-Memory Database and Oracle NoSQLDB. He has a B.Tech in Computer Science from IIT, Kharagpur, and an MS in Electrical Engineering from Stanford University.

Tuesday May 14th, 2024 9:45-10:15

Coffee Break - [In Speys]

Tuesday May 14th, 2024 10:15-11:45

AI for Databases I - [In Theater 12, Chair: Yong Zhang]

- DeepMapping: Learned Data Mapping for Lossless Compression and Efficient Lookup
  Lixi Zhou (Arizona State University); K. Selçuk Candan (Arizona State University); Jia Zou (Arizona State University)
- PURPLE: Making a Large Language Model a Better SQL Writer
  Tonghui Ren (Fudan University); Yuankai Fan (Fudan University); Zhenying He (Fudan University); Ren Huang (Fudan University); Jiaqi Dai (Fudan University); Can Huang (Fudan University); Yinan Jing (Fudan University); Kai Zhang (Fudan University); Yifan Yang (Transwarp Technology); X. Sean Wang (Fudan University)
- LeaderKV: Improving Read Performance of KV Stores via Learned Index and Decoupled KV Table
  Yi Wang (Shenzhen University); Jianan Yuan (Shenzhen University); Shangyu Wu (City University of Hong Kong`); Huan Liu (Shenzhen University); Jiaxian Chen (Shenzhen University); Chenlin Ma (Shenzhen University); Jianbin Qin (Shenzhen Institute of Computing Sciences, Shenzhen University)
- TRAP: Tailored Robustness Assessment for Index Advisors via Adversarial Perturbation
  Wei Zhou (Xiamen University); Chen Lin (Xiamen University); Xuanhe Zhou (Tsinghua); Guoliang Li (Tsinghua University); Tianqing Wang (Huawei)
- Duet: efficient and scalable hybriD neUral rElation undersTanding
  Kaixin Zhang (Harbin Institute of Technology); Hongzhi Wang (Harbin Institute of Technology); Yabin Lu (Harbin Institute of Technology); Ziqi Li (Harbin Institute of Technology); Chang Shu (Harbin Institute of Technology); Yu Yan (Harbin Institute of Technology); Donghua Yang (Harbin Institute of Technology)

Query Processing, Indexing, and Optimization I - [In Theater 11, Chair: Mohamed Eltabakh]

- Adaptive Recursive Query Optimization
  Anna Herlihy (EPFL); Guillaume Martres (EPFL); Anastasia Ailamaki (EPFL); Martin Odersky (EPFL)
- Ontology-Mediated Query Answering Using Graph Patterns with Conditions
  Ping Lu (Beihang Univ.); Ting Deng (" Beihang University, China"); Haoyuan Zhang (Beihang University); Yufeng Jin (Beihang University); Liu Feiyi (Beihang University); Tiancheng Mao (Beihang University); Lexiao Liu (Beihang University)
- An Efficient Algorithm for Continuous Complex Event Matching Using Bit-Parallelism
  Tao Qiu (Shenyang Aerospace University); Shenwang Jiang (Shenyang Aerospace University); Xiaochun Yang (Northeastern University); Bin Wang (Northeastern University); Chuanyu Zong (Shenyang Aerospace University); Rui Zhu (Shenyang Aerospace University)
- Personalized PageRanks over Dynamic Graphs -- The Case for Optimizing Quality of Service
  Zulun Zhu (Nanyang Technological University); Siqiang Luo (Nanyang Technological University); Wenqing Lin (Tencent); Sibo Wang (The Chinese University of Hong Kong); Dingheng Mo (Nanyang Technological University); Chunbo Li (Nanyang Technological University)
- PyTond: Efficient Python Data Science on the Shoulders of Databases
  Hesam Shahrokhi (University of Edinburgh); Amirali Kaboli (University of Edinburgh); Mahdi Ghorbani (University of Edinburgh); Amir Shaikhha (University of Edinburgh)

Graphs, Networks, and Semistructured Data I - [In Theater 10, Chair: Da Yan]

- Butterfly Counting over Bipartite Graphs with Local Differential Privacy
  Yizhang He (The University of New South Wales); Kai Wang (Shanghai Jiao Tong University); Wenjie Zhang (University of New South Wales); Xuemin Lin (Shanghai Jiaotong University); Wei Ni (CSIRO); Ying Zhang (University of Technology Sydney)
- Temporal Graph Generation Featuring Time-bound Communities
  Shuwen Zheng (Tsinghua University); Chaokun Wang (Tsinghua University); Cheng Wu (Tsinghua University); Yunkai Lou (Tsinghua University); Hao Feng (Tsinghua University); Xuran Yang (Tsinghua University)
- Breaking the Entanglement of Homophily and Heterophily in Semi-supervised Node Classification
  Henan Sun (Beijing Institute of Technology); Xunkai Li (Beijing Institute of Technology); Zhengyu Wu (Beijing Institute of Technology); Daohan Su (Beijing Institute of Technology); Ronghua Li (Beijing Institute of Technology); Guoren Wang (Beijing Institute of Technology)
- Efficient Core Decomposition over Large Heterogeneous Information Networks
  Yucan Guo (The Chinese University of Hong Kong, Shenzhen); Chenhao Ma (The Chinese University of Hong Kong, Shenzhen); Yixiang Fang (The Chinese University of Hong Kong, Shenzhen)
- Accelerating SpMV for Scale-Free Graphs with Optimized Bins
  YuAng Chen (The Chinese University of Hong Kong); Jeffrey Xu Yu (Chinese University of Hong Kong)

Data Mining and Knowledge Discovery I - [In Theater 9, Chair: Saeid Hosseini]

- Are ID Embeddings Necessary? Whitening Pre-trained Text Embeddings for Effective Sequential Recommendation
  lingzi zhang (Nanyang Technological University); Xin Zhou (Nanyang Technological University); Zhiwei Zeng (Nanyang Technological University); Zhiqi Shen (NTU)
- Structure- and Logic-aware Heterogeneous Graph Learning for Recommendation
  Anchen Li (Jilin University); Bo Yang (Jilin University); Huan Huo (University of Technology Sydney); Farookh Hussain (University of Technology Sydney); Guandong Xu (University of Technology Sydney, Australia)
- Robust Graph Augmentation for Recommendation
  Qianru Zhang (The University of Hong Kong); Lianghao Xia (University of Hong Kong); Xuheng Cai (The University of Hong Kong); Siu Ming Yiu (The University of Hong Kong); Chao Huang (University of Hong Kong); Christian S. Jensen (Aalborg University)
- From Chaos to Clarity: Anomaly Detection in Astronomical Observations
  Xinli Hao (Renmin University of China); Yile Chen (Nanyang Technological University); Chen Yang (China National Clearing Center); ZHIHUI DU (New Jersey Institute of Technology); Chaohong Ma (Renmin University of China); Chao Wu (National Astronomical Observations,Chinese Academy of Sciences); Xiaofeng Meng (Renmin University of China)
- Enhancing Topic Interpretability for Neural Topic Modeling through Topic-wise Contrastive Learning
  Xin Gao (Peking University); Yang Lin (Peking University); Ruiqing Li (Peking University); Yasha Wang (Peking University); Xu Chu (Tsinghua University); Xinyu Ma (Peking University); Hailong Yu (Peking University)

Very Large Data Science Applications/pipelines - [In Theater 4, Chair: Torben Bach Pedersen]

- SPES: Towards Optimizing Performance-Resource Trade-Off for Serverless Functions
  Cheryl Lee (The Chinese University of Hong Kong); Tianyi Yang (The Chinese University of Hong Kong); Zhouruixing Zhu (The Chinese University of Hong Kong, Shenzhen); Yintong Huo (Chinese University of Hong Kong); Pinjia He (The Chinese University of Hong Kong, Shenzhen); Yuxin Su (Sun Yat-sen University); Michael R Lyu (The Chinese University of Hong Kong)
- KGLiDS: A Platform for Semantic Abstraction, Linking, and Automation of Data Science
  Mossad Helali (Concordia University); Niki Monjazeb (Concordia University); Shubham Vashisth (Concordia University); Philippe Carrier (Concordia); Ahmed Helal (Concordia University); Antonio Cavalcante Araujo Neto (Borealis AI); Khaled Ammar (BorealisAI); Katja Hose (TU Wien); Essam Mansour (Concordia University)
- Efficiently Estimating Mutual Information Between Attributes Across Tables
  Aécio Santos (New York University); Flip Korn (Google, USA); Juliana Freire (New York University)
- Effective Entry-wise Flow for Molecule Generation
  Qifan Zhang (East China Normal University); Junjie Yao (East China Normal University); yuquan yang (East China Normal University); Yizhou Shi (University of Massachusetts); Wei Gao (University of Michigan); Xiaoling Wang (East China Normal University)
- HYPPO: Using Equivalences to Optimize Pipelines in Exploratory Machine Learning
  A. I. KONTAXAKIS (Université libre de Bruxelles); Dimitris Sacharidis (ULB); Alkis Simitsis (Athena Research Center); Alberto Abello (Universitat Politecnica de Catalunya); Sergi Nadal (Universitat Politècnica de Catalunya)

Information Integration and Data Quality I - [In Theater 3, Chair: Gerardo Vitagliano]

- FedMix: Boosting with Data Mixture for Vertical Federated Learning
  Yihang Cheng (University of Science and Technology of China); Lan Zhang (University of Science and Technology of China); JunYang Wang (University of Science and Technology of China); Xiaokai Chu (Institute of Computing Technology, Chinese Academy of Sciences); Dongbo Huang (Tencent); Lan Xu (Tencent)
- DMRNet: Effective Network for Accurate Discharge Medication Recommendation
  Shi jiyun (Beijing Institute of Technology); Yuqiao Wang (Beijing Institute of Technology); Chi Zhang (Beijing Institute of Technology); Zhaojing Luo (Beijing Institute of Technology); Chengliang Chai (Beijing Institute of Technology); Meihui Zhang (Beijing Institute of Technology)
- BClean: A Bayesian Data Cleaning System
  Yaoshu Wang (Shenzhen Institute of Computing Sciences, Shenzhen University); Jianbin Qin (Shenzhen Institute of Computing Sciences, Shenzhen University); Sifan Huang (Shenzhen Institute of Computing Sciences, Shenzhen University); Jing Zhu (Shenzhen University); Yifan Zhang (Shenzhen Institute of Computing Sciences, Shenzhen University); Yukai Miao (University of New South Wales); Rui Mao (Shenzhen University); Makoto Onizuka (Osaka University); Chuan Xiao (Osaka University, Nagoya University)
- MultiEM: Efficient and Effective Unsupervised Multi-Table Entity Matching
  Xiaocan Zeng (Zhejiang University); Pengfei Wang (Zhejiang University); Yuren Mao (Zhejiang University); Lu Chen (Zhejiang University); Xiaoze Liu (ZJU); Yunjun Gao (Zhejiang University)
- A Critical Re-evaluation of Benchmark Datasets for (Deep) Learning-Based Matching Algorithms
  George Papadakis (University of Athens); Nishadi Kirielle (The Australian National University); Peter Christen (The Australian National University); Themis Palpanas (Université Paris Cité)

Industry - Cloud Database Systems - [In Theater 2, Chair: Renata Borovica-Gajic]

- Invited Talk: Vector Search and Databases
  Yannis Papakonstantinou (Google)
- GaussDB-Global: A Geographically Distributed Database System
  Puya Memarzia (Huawei); Huaxin Zhang (Huawei Canada Ltd.); Kelvin Ho (Huawei Canada Technologies Inc.); Ronen Grosman (Huawei); Jiang Wang (Huawei)
- Towards a Shared-storage-based Serverless Database Achieving Seamless Scale-up and Read Scale-out
  Yingqiang Zhang (Alibaba Group); xinjun Yang (Alibaba Group); Hao Chen (Alibaba Group); Feifei Li (Alibaba Group); jiawei xu (Alibaba Group); Jie Zhou (Alibaba); Xudong Wu (Alibaba Group); Qiang Zhang (Alibaba Cloud)
- Optimized Locking in SQL Azure
  Chaitanya Sreenivas Ravella (Microsoft); Prashanth Purnananda (Microsoft); Hanuma Kodavalla (Microsoft); Peter Byrne (Microsoft); Adrian-Leonard Radu (Microsoft); Wayne Chen (Microsoft); Srikanth Sampath (Microsoft); Naga Bhavana Atluri (Microsoft); Srinag Rao (Microsoft); Priyanka A Kakade (Microsoft)
- Separation Is for Better Reunion: Data Lake Storage at Huawei
  Xin Tang (University of Wisconsin-Madison); Chengliang Chai (Beijing Institute of Technology)*; Dawei Zhao (Huawei); Haohai Ma (Huawei); Yong Zheng (Huawei); Zhenyong Fan (Huawei); Xin Wu (Huawei); Jiaquan Zhang (HW); Rui Zhang (HW); Duanshun Li (HW); Yi He (HW); Keji Huang (Huawei); Guangbin Meng (Huawei); Yidong Wang (Huawei); Yuefeng Zhou (Huawei); Tao Tao (China Mobile); Lirong Jian (Hashdata); Jiwu Shu (Tsinghua); Yu-Ping Wang (Beijing Institute of Technology); Ye Yuan (Beijing Institute of Technology); Guoren Wang (Beijing Institute of Technology); Guoliang Li (Tsinghua University)

Tutorial on Bipartite Graph Analytics: Current Techniques and Future Trends [In Theater 1]

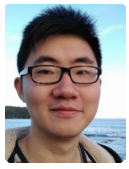
Abstract:
As the field of data science continues to evolve, bipartite graphs have emerged as a fundamental structure in numerous applications, drawing significant interest from both academic and industrial communities. Bipartite graphs are a specific type of graph consisting of two distinct sets of vertices, where connections only occur between vertices of different sets. Examples include e-commerce networks and biological networks. Analytics of bipartite graphs has become an important research topic in the era of big data. This tutorial aims to shed light on analysis methods for bipartite graphs, categorizing them into three areas: classical models, learning-based models, and application-driven models. We start by outlining the importance of bipartite graph analytics, and the unique challenges that need to be addressed. Then, we conduct a thorough review of existing works on bipartite graph analytics. We also compare and analyze the models and solutions in these works. Finally, we point out new research directions.

Presenters:

Ying Zhang is a Professor and ARC Future Fellow (2017- 2021) at Australia Artificial Intelligence Institute (AAII), the University of Technology Sydney (UTS). He received his BSc and MSc degrees in Computer Science from Peking University, and PhD in Computer Science from the University of New South Wales. His research interests include query processing and analytics on large-scale data with focus on graphs and high dimensional data.

Kai Wang Kai Wang is an assistant professor at Antai College of Economics and Management, Shanghai Jiao Tong University. He received the BEng degree in Computer Science from Zhejiang University in 2016, and the PhD degree in Computer Science from the University of New South Wales in 2020. His research interests lie in big data analytics, especially for the graph/network and spatial data.

Wenjie Zhang is a full Professor and ARC Future Fellow in School of Computer Science and Engineering, the University of New South Wales, Australia. She has published over 200 research papers in top venues in database area such as SIGMOD, VLDB, ICDE, PODS, TODS, VLDBJ, and TKDE. Her papers were nominated as Best of SIGMOD and ICDE, and receive Best Paper Awards from DASFAA, WISE, APWeb and ADC. She received Chris Wallace Award from Australasian Computing Research and Education (CORE) in 2019. She serves as an Associate Editor for TKDE and VLDB Journal, and area chair for ICDE/VLDB/ICDM.



Hanchen Wang is a postdoctoral research associate at AAII, FEIT, University of Technology Sydney, Australia. He received BSc degree from Zhejiang University, China in 2016, and PhD degree from University of Technology Sydney, Australia in 2021. His research interests include graph analytics and machine learning for databases.

Poster Session - [In Speys]

Posters of the 15:45 Session

Poster Session - [In B2B]

Posters of the 15:45 Session

ORACLE

Goodbye, database complexity. Hello, simplicity.

Reduce complexity and cost with the one database for all your workloads. Cloud native, converged, and automated, Oracle Autonomous Database eliminates manual maintenance hassles and delivers top performance, security, and automatic scaling for your changing needs. It also supports the full development lifecycle for modern applications.

Get simplicity in the cloud or on-premises.

Oracle Autonomous Database
Learn more at
oracle.com/autonomous-database

Tuesday May 14th, 2024 11:45-13:15

Lunch - [In Speys]

Tuesday May 14th, 2024 13:15-15:15

Sponsor Talks [In Theater 12, Chair:TBD]

- TBD
  by Tirthankar Lahiri (Senior Vice President of Data, Oracle)

  Tirthankar Lahiri is Senior Vice President of Data, In-Memory and AI Vector technologies at Oracle. He is responsible for the Oracle Database Data Engine as well as the TimesTen In-Memory Database and Oracle NoSQLDB. He has a B.Tech in Computer Science from IIT, Kharagpur, and an MS in Electrical Engineering from Stanford University.

- Alibaba Cloud ApsarasDB:Towards a Serverless and AI-Driven One-Stop Data Platform
  by Yingqiang Zhang (Senior Staff Engineer, Alibaba Cloud)

  Alibaba Cloud offers one of the most extensive portfolios of cloud database solutions and can provide the necessary solutions to store, process, analyze and manage your data to support and add value to your business. Our database systems support all the mainstream open-source and commercial database solutions, including MongoDB, PostgreSQL, MySQL, SQL Server, and Redis. Alibaba Cloud has been named a Leader in the 2023 Gartner® Magic Quadrant™ for Cloud Database Management Systems for the 4th consecutive year.

  Yingqiang Zhang is the Senior Staff Engineer and Director at Alibaba Cloud, leading several R&D teams within the PolarDB department and focusing on the next-generation architecture of cloud-native databases. His teams have designed and implemented critical features for PolarDB, such as strongly consistent reads on followers (presented at VLDB'23) and a serverless model for PolarDB (best industry paper at ICDE'24). Additionally, Zhang and his teams explored distributed shared-memory for cloud-native databases (presented at VLDB'21 and ASPLOS'23), and they have designed and implemented the PolarDB Multi-Primary architecture (to appear in SIGMOD'24). His research spans cloud-native database architectures, distributed systems, and hardware co-design. He has published papers in several top conferences such as SIGMOD, VLDB, ICDE, and ASPLOS. He also serves for ICDE 2024 and ChinaSys Workshop 2024 as a PC member.

- Hybrid Data Analytical Processing and Hardware Acceleration Research at ByteDance
  by Jianjun Chen (Director of Infrastructure System Lab, ByteDance US)

  Dr Jianjun Chen is Director of Infrastructure System Lab at ByteDance US, where he leads a team of top-notch researchers and software engineers to work on cutting edge technologies in the infrastructure system related areas, including but not limited to database, storage, computing, networking, ML, hardware/software codesign etc. Before that, he was a technical VP of Huawei US Silicon Valley R&D Center, leading advanced database research and development in the Huawei database group. Dr. Chen received his Ph.D in 2001 from the Computer Sciences department of University of Wisconsin, Madison and is a recipient of the SIGMOD 10 year Test-Time award in 2010 for his visionary work in scalable continuous query processing, as part of his Ph.D dissertation research.

- Huawei Cloud GaussDB, a Better Way to Database
  by Nikos Ntarmos (Director, Huawei)

  Nikos Ntarmos is the Director of the Database Lab at Huawei's Edinburgh Research Centre, where he leads a team of researchers and engineers working on designing and implementing next generation database management systems, for environments ranging from small routers/switches to big-iron servers and the cloud. His research interests lie in the areas of distributed computing and (large-scale) data management systems, with a focus on issues pertaining to storage, indexing and query processing and optimization, in embedded databases, distributed data stores, multi-model databases, geo-distributed data management infrastructures, and joint at-rest/streaming data processing systems. He received his PhD in computer engineering and informatics from the University of Patras in 2008. He is a member of the IEEE and the ACM, and a Fellow of the UK Higher Education Academy. Before joining Huawei, he held academic posts at the University of Glasgow and the University of Ioannina, and research fellow posts at the University of Glasgow and the University of Patras. He has also worked as a research engineer, software engineer, and senior systems administrator, and routinely contributes to open-source projects. He has supervised or co-supervised 7 PhD and over 50 MSc students, was a recipient of the best paper award at the ACM CIKM 2006 conference and the best student paper award at the IEEE Big Data 2018 conference, and has served in the Program Committees of multiple international conferences (including ACM SIGMOD, IEEE ICDE, ACM/IFIP/USENIX Middleware, WISE, IEEE CCNC, etc.).

- OceanBase: From OLTP to HTAP
  by Chuanhui Yang (CTO, OceanBase)

  OceanBase is a distributed SQL database built from scratch from 2010 by Ant Group. Initially, it was used by all mission critical systems in Alipay and then extended to more than 1000 external customers outside Alipay. It started as a distributed OLTP database, but more and more customers used it in mixed workload scenarios such as OLAP and NOSQL. This talk will mainly focus on the technical challenges from distributed OLTP to HTAP, such as column storage, resource isolation, complex queries, multi-model in distributed SQL database.

  Chuanhui Yang, also known by his nickname "Rizhao," serves as the Chief Technology Officer for OceanBase, a native distributed database developed by Ant Group. His areas of expertise include large-scale cloud computing systems and distributed databases. In 2010, he joined OceanBase as one of its founding members. He has played a pivotal role in every phase of OceanBase's architecture design and technology development, leading to its successful implementation across Ant Group. Additionally, Yang led the team that twice broke world records in the TPC-C benchmarks. He is the author of "Large-scale Distributed Storage Systems: Theory Analysis and Practical Framework" and contributed to "Analysis of OceanBase Database Source Code." Currently, Yang Chuanhui and his team at OceanBase are dedicated to creating a next-generation enterprise-level distributed database that is more open, flexible, efficient, and user-friendly.

- Engineering at Snowflake in Berlin; Past, Present, and Future
  by Martin Schoenert (Software Engineering Manager, Snowflake)

- Research and Applications of Tencent Database Systems
  by Yuxing Chen, (Senior Engineer, Tencent)

  Yuxing Chen received his Ph.D. in computer science from the University of Helsinki, Finland, 2021, He currently works as a senior research engineer in the database R&D department at Tencent, China. His research interests focus on database performance and evaluation, transaction processing, and database storage. He has published papers in several prestigious conferences and journals such as SIGMOD, VLDB, ICDE, and TKDE.

Tuesday May 14th, 2024 15:15-15:45

Coffee Break - [In Speys]

Tuesday May 14th, 2024 15:45-16:21

AI for Databases II - [In Theater 12, Chair: Arijit Khan]

- Knowledge Graph Enhanced Multimodal Transformer for Image-Text Retrieval
  Juncheng Zheng (Beijing University of Posts and Telecommunications); Meiyu Liang (Beijing University of Posts and Telecommunications); Yang Yu (Beijing University of Posts and Telecommunications); Yawen Li (Beijing University of Posts and Telecommunications); Zhe Xue (Beijing University of Posts and Telecommunications)
- Functionality-Aware Database Tuning via Multi-Task Learning
  Zhongwei Yue (East China Normal University | ECNU · Department of Computer Science & Technology); Shujian Peng (East China Normal University); Peng Cai (East China Normal University); Xuan Zhou (East China Normal University); Huiqi Hu (East China Normal University); Rong Zhang (East China Normal University); Quanqing Xu (OceanBase, Ant Group); Chuanhui Yang (OceanBase)

Query Processing, Indexing, and Optimization II - [In Theater 11, Chair: Ashwin Lall]

- Efficient Fault Tolerance for Pipelined Query Engines via Write-ahead Lineage
  Ziheng Wang (Stanford); Alex Aiken (Stanford University)
- Independent Range Sampling on Interval Data
  Daichi Amagata (Osaka University)

Graphs, Networks, and Semistructured Data II - [In Theater 10, Chair: Yuchen Li]

- PlatoD2GL: An Efficient Dynamic Deep Graph Learning System for Graph Neural Network Training on Billion-Scale Graphs
  Xing Huang (Tencent); Dandan LIN (Shenzhen Institute of Computing Sciences); Weiyi Huang (Tencent); Shijie Sun (Tencent Inc.); Jie Wen (Tencent Group Limit); Chuan Chen (Tencent)
- Quantum Algorithms for the Maximum K-Plex Problem
  Xiaofan Li (Nanyang Technological University); Gao Cong (Nanyang Technological Univesity); Rui Zhou (Swinburne University of Technology)

Data Mining and Knowledge Discovery II - [In Theater 9, Chair: Zeyi Wen]

- Improve ROI with Causal Learning and Conformal Prediction
  Meng Ai (China Mobile Information Technology Co., Ltd); Zhuo Chen (China Mobile); Jibin Wang (China Mobile Information Technology Center); Jing Shang (China Mobile Information Technology Co. Ltd.); Tao Tao (China Mobile Information Technology Co. Ltd.); Zhen Li (China Mobile Information Technology Co., Ltd)
- Hide Your Model: A Parameter Transmission-free Federated Recommender System
  Wei Yuan (The University of Queensland); Chaoqun Yang (Griffith University); Liang Qu (The University of Queensland); Quoc Viet Hung Nguyen (Griffith University); Jianxin Li (Deakin University); Hongzhi Yin (The University of Queensland)

Spatial Databases and Temporal Databases I - [In Theater 4, Chair: Harry Kai-Ho Chan]

- TrendSharing: a Framework to Discover and Follow the Trends for Shared Mobility Services
  Zhan Jiexi (East China Normal University); Han Wu (ECNU); Peng Cheng (East China Normal University); Libin Zheng (Sun Yat-sen University); Lei Chen (Hong Kong University of Science and Technology); Chen Zhang (The Hong Kong Polytechnic University); Xuemin Lin (Shanghai Jiaotong University); Wenjie Zhang (University of New South Wales)
- Collectively Simplifying Trajectories in a Database: A Query Accuracy Driven Approach
  ZHENG WANG (Huawei Singapore Research Center); Cheng Long (Nanyang Technological University); Gao Cong (Nanyang Technological Univesity); Christian S. Jensen (Aalborg University)

Information Integration and Data Quality II - [In Theater 3, Chair: Reynold Cheng]

- Online Query-based Data Pricing with Time-discounting Valuations
  Yicheng Fu (Zhejiang university); Xiaoye Miao (Zhejiang University); Huanhuan Peng (Zhejiang University); Chongning Na (Zhejiang Lab); Shuiguang Deng (Zhejiang University); Jianwei Yin (Zhejiang University)
- Representation Learning for Entity Alignment in Knowledge Graph: A Design Space Exploration
  Peng Huang (Beijing Institute of Technology); Meihui Zhang (Beijing Institute of Technology); Ziyue Zhong (Beijing Institute of Technology); Chengliang Chai (Beijing Institute of Technology); Ju Fan (Renmin University of China)

Data Mining and Knowledge Discovery IV - [In Theater 2, Chair: Jianbin Qin]

- Efficient Set-based Order Dependency Discovery with a Level-wise Hybrid Strategy
  Yihan Li (Fudan University); Ruifeng Li (Fudan University); Zijing Tan (Fudan University); Shuai Ma (Beihang University)
- Meta-optimized Joint Generative and Contrastive Learning for Sequential Recommendation
  Yongjing Hao (Soochow University); Pengpeng Zhao (Soochow University); Junhua Fang (Soochow University); JIanfeng Qu (Soochow University); Guanfeng Liu (Macquarie University); Fuzhen Zhuang (Institute of Artificial Intelligence, Beihang University); Victor S. Sheng (Texas Tech University); Xiaofang Zhou (The Hong Kong University of Science and Technology)

Data Mining and Knowledge Discovery V - [In Theater 1, Chair: Jiawei Jiang]

- Multi-modal Siamese Network for Few-shot Knowledge Graph Completion
  Yuyang Wei (School of Computer Science and Technology, Soochow University); Wei Chen (Soochow University); Xiao-Fang Zhang (Soochow University); Pengpeng Zhao (Soochow University); JIanfeng Qu (Soochow University); Lei Zhao (Soochow University)
- Local-Global History-aware Contrastive Learning for Temporal Knowledge Graph Reasoning
  Wei Chen (Beijing Jiaotong University); Huaiyu Wan (Beijing Jiaotong University); Yuting Wu (Beijing Jiaotong University); Jiayaqi Cheng (Beijing jiaotong university); Shuyuan Zhao (Beijing Jiaotong University); Yuxin Li (BJTU); Youfang Lin (Beijing Jiaotong University)

Poster Session - [In Speys]

Posters of the 10:30 Session

Demonstrations - Group A - [In B2B]

- Entity/Relationship Profiling
  by Henning Koehler (Massey University), Sebastian Link (University of Auckland)
- GA-Tags: Data Enrichment with an Automatic Tagging System Utilizing Large Language Models
  by Genki Kusano (NEC)
- Comparing Personalized Relevance Algorithms for Directed Graphs
  by Luca Cavalcanti (University of Trento), Cristian Consonni (Joint Research Centre, European Commission), Martin MB Brugnara (University of Trento), David Laniado (Eurecat), Alberto Montresor (University of Trento)
- FSM-Explorer: An Interactive Tool for Frequent Subgraph Pattern Mining from a Big Graph
  by Jalal Khalil (St. Cloud State University), Da Yan (Indiana University Bloomington), Lyuheng Yuan (Indiana University Bloomington), Jiao Han (George Washington University), Saugat Adhikari (Indiana University Bloomington), Cheng Long (Nanyang Technological University), Yang Zhou (Auburn University)
- TASKS: A Real-Time Query System for Instant Error-Tolerant Spatial Keyword Queries on Road Networks
  by Chengyang Luo (Zhejiang University), lu jin (Zhejiang University), Qing Liu (Zhejiang University), Yunjun Gao (Zhejiang University), Lu Chen (Zhejiang University)
- VASIM: Vertical Autoscaling Simulator Toolkit
  by Anna Pavlenko (Microsoft Gray Systems Lab), Karla Saur (Microsoft), Yiwen Zhu (Microsoft), Brian Kroth (Microsoft), Joyce Cahoon (Microsoft), Jesús Camacho-Rodríguez (Microsoft)
- Demonstration of FeVisQA: Free-Form Question Answering over Data Visualization
  by Yuanfeng Song (The Hong Kong University of Science and Technology), 锦伟 卢 (深圳大学), Xuefang Zhao (WeBank Co., Ltd), Raymond Chi-Wing Wong (Hong Kong University of Science and Technology), Haodi Zhang (Shenzhen University)
- CleanEr: Interactive, Query-Guided Error Mitigation for Data Cleaning Systems
  by Ran Schreiber (Bar-Ilan University), Yael Amsterdamer (Bar-Ilan university)
- Wearables for Health (W4H) Toolkit for Acquisition, Storage, Analysis and Visualization of Data from Various Wearable Devices
  by Arash Hajisafi (University of Southern California), Maria Despoina Siampou (University of Southern California), Jize Bi (University of Southern California), Luciano Nocera (University of Southern California), Cyrus Shahabi (Computer Science Department. University of Southern California)

Demonstrations - Group B - [In B2B]

- Chat2Query: A Zero-Shot Automatic Exploratory Data Analysis System with Large Language Models
  by Jun-Peng Zhu (East China Normal University), Peng Cai (East China Normal University), Boyan Niu (PingCAP), Zheming Ni (PingCAP), Kai Xu (PingCAP), Jiajun Huang (PingCAP), Jianwei Wan (PingCAP), Shengbo Ma (PingCAP), Bing Wang (PingCAP), Donghui Zhang (PingCAP), Liu Tang (PingCAP), Qi Liu (PingCAP)
- EADS: An Early Anomaly Detection System for Sensor-based Multivariate Time Series
  by Yihao Ang (National University of Singapore), Qiang Huang (National University of Singapore), Anthony K. H. Tung (NUS), Zhiyong Huang (NUS School of Computing)
- d_{symb} playground: an interactive tool to explore large multivariate time series datasets
  by Sylvain W Combettes (Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli), Paul Boniol (Université de Paris), Charles Truong (Université Paris-Saclay, ENS Paris-Saclay), Laurent Oudre (ENS Paris-Saclay)
- ADecimo: Model Selection for Time Series Anomaly Detection
  by Paul Boniol (Université de Paris), Emmanouil Sylligardos (FORTH), John Paparrizos (The Ohio State University), Panos Trahanias (FORTH), Themis Palpanas (Université Paris Cité)
- ChatGraph: Chat with Your Graphs
  by Yun PENG (Guangzhou University), Sen Lin (Guangzhou University), Qian Chen (Hong Kong Baptist University), Lyu Xu (Hong Kong Baptist University), Xiaojun Ren (Guangzhou University), Yafei Li (Zhengzhou University), Jianliang Xu (Hong Kong Baptist University)
- RLQDAG: a Fast Plan Enumerator for Top-down Recursive Query Optimizers
  by Amela Fejza (Inria), Pierre Genevès (CNRS), Nabil Layaïda (Inria)
- KGSEC: A Modular Framework for Knowledge Graph Schema Extraction and Comparison
  by Petros Skoufis (Athena Research Center), Dimitrios Skoutas (Athena Research Center)
- QFusor: A UDF Optimizer Plugin for SQL Databases
  by Konstantinos Chasialis (Athena RC), Theoni Palaiologou (Athena Research Center), Yannis E Foufoulas (University of Athens), Alkis Simitsis (Athena Research Center), Yannis Ioannidis (University of Athens)
- ARTS: A System for Aggregate Related Table Search
  by Junjie Xing (University of Michigan), H. V. Jagadish (University of Michigan)

Poster Session - [In B2B]

Posters of the PhD Symposium Session

World-leading Databases Technologies and Products

Alibaba Cloud

ApsaraDB

Named as a LEADER in Gartner MQ for Cloud DBMS(2020-2023)

Alibaba Cloud Database Market Share

- 1st In China (IDC 2020-2022)
- 1st In Asia Pacific (Gartner 2021)
- 4st Cloud Database Provider Worldwide (Gartner 2022)

150,000+ Enterprise customers

700,000+ Database instances migrated

About PolarDB

PolarDB is a cloud-native relational database service developed in-house by Alibaba Cloud. It is fully compatible with MySQL and PostgreSQL, and is highly compatible with Oracle syntax. With its cloud-native architectures (integrated or distributed), storage-compute decoupling, and software-hardware co-design, PolarDB delivers unparalleled elasticity, performance, storage capability, availability, and cost-efficiency.

Engine	PolarDB for MySQL	PolarDB for PostgreSQl	PolarDB for Xscale (PolarDB-X)
Ecosystem	MySQL	PostgreSQL and Oracle	MySQL
Description	• A cloud-native relationa ldatabase • 100% compatibility with MySQL • CPU-memory-storage decoupled • Software and hardware co-design	• 100% compatibility with PostgreSQL • High compatibility with Oracle syntax • Deliver capabilities on-par-with commercial databases • Advantages in availability reliability, linear scalability cost-efficiency, and performance	• Distributed architecture • Integration of distributed SQL engine and self-developed storage • Scenarios with massive data, high concurrency and throughput

Tuesday May 14th, 2024 16:21-18:00

AI for Databases III - [In Theater 12, Chair: Wei Huang]

- COSTREAM: Learned Cost Model for Operator Placement in Edge-Cloud Environments
  Roman Heinrich (DHBW Mannheim); Carsten Binnig (TU Darmstadt); Harald Kornmayer (DHBW Mannheim); Manisha Luthra (TU Darmstadt and DFKI)
- SiloFuse: Cross-silo Synthetic Data Generation with Latent Tabular Diffusion Models
  Aditya Shankar (TU Delft); Rihan Hai (TU Delft); Hans Brouwer (BlueGen.ai); Lydia Chen (TU Delft)
- Explainable Database Management System Configuration Tuning through Counterfactuals
  Xinyue Shao (Harbin Institute of Technology); Hongzhi Wang (Harbin Institute of Technology); Xiao Zhu (Harbin Institute of Technology); TianYu Mu (Harbin Institute of Technology); Yan Zhang (Harbin Institute of Technology)
- In Situ Neural Relational Schema Matcher
  xingyu du (zhejiang university); Gongsheng Yuan (Zhejiang University); Sai Wu (Zhejiang University); Gang Chen (Zhejiang University); Peng Lu (Institute of Computing Innovation, Zhejiang University)
- Towards Exploratory Query Optimization for Template-based SQL Workloads
  jieming feng (northwestern ploytechnical university); Zhanhuai Li (Northwestern Polytechnical University); Qun Chen (Northwestern Polytechnical University)

Query Processing, Indexing, and Optimization III - [In Theater 11, Chair: Maximilian E Schüle]

- Incremental Fusion: Unifying Compiled and Vectorized Query Execution
  Benjamin Wagner (Technical University of Munich); André Kohn (Technical University of Munich); Peter Boncz (Centrum Wiskunde & Informatica); Viktor Leis (Technische Universität München)
- The Indistinguishability Query
  Ashwin Lall (Denison University)
- Range Cache: An Efficient Cache Component for Accelerating Range Queries on LSM-Based Key-Value Stores
  Xiaoliang Wang (University of Science and Technology of China); Peiquan Jin (University of Science and Technology of China); Yongping Luo (University of Science and Technology of China); Zhaole Chu (University of Science and Technology of China)
- Optimizing Context-Enhanced Relational Joins
  Viktor Sanca (EPFL); Manos Chatzakis (EPFL); Anastasia Ailamaki (EPFL)
- IndeXY: A Framework for Constructing Indexes Larger than Memory
  Chen Zhong (University of Texas at Arlington); Qingqing Zhou (DB365000); Yuxing Chen (Tencent); Xingsheng Zhao (University of Texas at Arlington); Kuang He (Tencent); Anqun Pan (Tencent Inc., China); Song Jiang (University of Texas, Arlington)

Graphs, Networks, and Semistructured Data III - [In Theater 10, Chair: Chenhao Ma]

- Fast Iterative Graph Computing with Updated Neighbor States
  Yijie Zhou (Northeastern University); Shufeng Gong (NorthEastern University); Feng Yao (Northeastern University); Hanzhang Chen (Northeastern University); Song Yu (Northeastern University); Pengxi Liu (Northeastern University); Yanfeng Zhang (Northeastern University); Ge Yu (Northeastern University); Jeffrey Xu Yu (Chinese University of Hong Kong)
- Querying Numeric-Constrained Shortest Distances on Road Networks
  Mingyu Yang (The Hong Kong University of Science and Technology (Guangzhou)); Wentao Li (The Hong Kong University of Science and Technology (Guangzhou)); Wei Wang (Hong Kong University of Science and Technology (Guangzhou)); Dong Wen (University of New South Wales); Lu Qin (UTS)
- Mining Quasi-Periodic Communities in Temporal Network
  Yue Zeng (Beijing Institute of Technology); Hongchao Qin (Beijing Institute of Technology); Ronghua Li (Beijing Institute of Technology); Kai Wang (Shanghai Jiao Tong University); Guoren Wang (Beijing Institute of Technology); Xuemin Lin (Shanghai Jiao Tong University)
- GraphRARE: Reinforcement Learning Enhanced Graph Neural Network with Relative Entropy
  Tianhao Peng (Beihang University); Wenjun Wu (Beihang University); Haitao Yuan (Nanyang Technological University); Zhifeng Bao (RMIT University); zhao pengrui (BUAA); Xin Yu (Beihang University); Xuetao Lin (Beihang University); Yu Liang (Beihang University); yanjun pu (beihang)
- Querying Historical Cohesive Subgraphs over Temporal Bipartite Graphs
  Shunyang Li (University of New South Wales); Kai Wang (Shanghai Jiao Tong University); Xuemin Lin (Shanghai Jiaotong University); Wenjie Zhang (University of New South Wales); Yizhang He (The University of New South Wales); Long Yuan (Nanjing University of Science and Technology)

Data Mining and Knowledge Discovery III - [In Theater 9, Chair: John Paparrizos]

- TimeDRL: Disentangled Representation Learning for Multivariate Time-Series
  Ching Chang (NYCU); ChiaoTung Chan (NYCU); Wei-Yao Wang (National Yang Ming Chiao Tung University); Wen-Chih Chris Peng (NYCU CS); Tien-Fu Chen (National Yang Ming Chiao Tung Univ)
- Boosting Meaningful Dependency Mining with Clustering and Covariance Analysis
  Xi Wang (National University of Defense Technology); Ruochun Jin (National University of Defense Technology); Wanrong Huang (College of Computer Science and Technology, National University of Defense Technology); Yuhua Tang (National University of Defense Technology)
- Uncovering the Limitations of Eliminating Selection Bias for Recommendation: Missing Mechanisms, Disentanglement, and Identifiability
  Haoxuan Li (Peking University); Shuyi Wang (University of Pennsylvania); Honglei Zhang (Beijing Jiaotong University); Chunyuan Zheng (University of California, San Diego); Xu Chen (Renmin University of China); Li Liu (Chongqing University); Shanshan Luo (Beijing Technology and Business University); Peng Wu (Beijing Technology and Business University)
- Scaling up Multivariate Time Series Pre-Training with Decoupled Spatial-Temporal Representations
  Rui Zha (University of Science and Technology of China); Le Zhang (Baidu Research); Shuangli Li (University of Science and Technology of China); jingbo zhou (Baidu Research); Tong Xu (University of Science and Technology of China); Enhong Chen (University of Science and Technology of China); Hui Xiong (Hong Kong University of Science and Tech)
- Meta-optimized Structural and Semantic Contrastive Learning for Graph Collaborative Filtering
  Yongjing Hao (Soochow University); Pengpeng Zhao (Soochow University); JIanfeng Qu (Soochow University); Lei Zhao (Soochow University); Guanfeng Liu (Macquarie University); Fuzhen Zhuang (Institute of Artificial Intelligence, Beihang University); Victor S. Sheng (Texas Tech University); Xiaofang Zhou (The Hong Kong University of Science and Technology)

Spatial Databases and Temporal Databases II - [In Theater 4, Chair: Sijie Ruan]

- Efficient Learning-based Top-k Representative Similar Subtrajectory Query
  Kunming Wang (Guangzhou university); Shiyu Yang (Guangzhou University); Jin Jiabao (East China Normal University); Peng Cheng (East China Normal University); Jianye Yang (Guangzhou University); Xuemin Lin (Shanghai Jiaotong University)
- Urban Region Representation Learning with Attentive Fusion
  Fengze Sun (University of Melbourne); Jianzhong Qi (The University of Melbourne); Yanchuan Chang (The University of Melbourne); Xiaoliang Fan (Xiamen University); Shanika Karunasekera (The University of Melbourne, Australia); Egemen Tanin (University of Melbourne)
- LightTR: A Lightweight Framework for Federated Trajectory Recovery
  Ziqiao Liu (UESTC); Hao Miao (Aalborg University); Yan Zhao (Aalborg University); Chenxi Liu (Nanyang Technological University); Kai Zheng (University of Electronic Science and Technology of China); Huan Li (Zhejiang University)
- Learning Time-aware Graph Structures for Spatially Correlated Time Series Forecasting
  Minbo Ma (Southwest Jiaotong University); Jilin Hu (Aalborg University); Christian S. Jensen (Aalborg University); Fei Teng (Southwest Jiaotong University); Peng Han (KAUST); Zhiqiang Xu (MBZUAI); Tianrui Li (School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, 611756, China)
- Deep Dirichlet Process Mixture Model for Non-parametric Trajectory Clustering
  Di Yao (Institute of Computing Technology, Chinese Academy of Sciences); Jin Wang (UCLA); Wenjie Chen (Institute of Computing Technology, Chinese Academy of Sciences); Fangda Guo (Institute of Computing Technology, Chinese Academy of Sciences); Peng Han (KAUST); Jingping Bi (Institute of Computing Technology, Chinese Academy of Sciences)

Information Integration and Data Quality III - [In Theater 3, Chair: Hazar Harmouch]

- Fairness-aware Data Preparation for Entity Matching
  Nima Shahbazi (University of Illinois at Chicago); Jin Wang (Megagon Labs); Zhengjie Miao (Simon Fraser University); Nikita Bhutani (Megagon Labs)
- Mitigating Data Sparsity in Integrated Data through Text Conceptualization
  Md Ataur Rahman (Universitat Politècnica de Catalunya); Sergi Nadal (Universitat Politècnica de Catalunya); Oscar Romero (Universitat Politècnica de Catalunya); Dimitris Sacharidis (ULB)
- Measuring Approximate Functional Dependencies: a Comparative Study
  Marcel Parciak (Hasselt University); Sebastiaan Weytjens (Hasselt University); Niel Hens (Hasselt University); Frank Neven (Hasselt University); Liesbet M. Peeters (Uhasselt); Stijn Vansummeren (Hasselt University)
- Efficient Relaxed Functional Dependency Discovery with Minimal Set Cover
  Xiaoou Ding (Harbin Institute of Technology); Yida Liu (Harbin Institute of Technology); Hongzhi Wang (Harbin Institute of Technology); Chen Wang (" Tsinghua University, China"); Yichen Song (HIT); Donghua Yang (Harbin Institute of Technology); Jianmin Wang ("Tsinghua University, China")
- Gen-T: Table Reclamation in Data Lakes
  Grace Fan (Northeastern University); Roee Shraga (WPI); Renée J. Miller (Northeastern University)

Industry - Machine Learning and Data Science - [In Theater 2, Chair: Essam Mansour]

- Deep Learning with Spatiotemporal Data: A Deep Dive into GeotorchAI
  Kanchan Chowdhury (Arizona State University); Mohamed Sarwat (Wherobots Inc.)
- DataLore: can LLM find all lost scrolls in a data repository? [Short Paper]
  Yuze Lou (University of Michigan); Chuan Lei (Amazon Web Services); Xiao Qin (Amazon); Zichen Wang (Amazon); Christos Faloutsos (Amazon); Rishita Anubhai (Amazon AWS AI); Huzefa Rangwala (George Mason University)
- ETUDE - Evaluating the Inference Latency of Session-Based Recommendation Models at Scale
  Barrie Kersbergen (bol.com); Olivier Sprangers (University of Amsterdam); Frank Kootte (bol.com); Shubha Guha (University of Amsterdam); Maarten de Rijke (University of Amsterdam); Sebastian Schelter (University of Amsterdam)
- CoachLM: Automatic Instruction Revisions Improve the Data Quality in LLM Instruction Tuning
  Yilun Liu (Huawei); Shimin Tao (Huawei); Xiaofeng Zhao (Huawei); Ming Zhu (Huawei); Wenbing Ma (Huawei); Junhao Zhu (Huawei); Chang Su (Huawei); Yutai Hou (Huawei); Miao Zhao (Huawei); Min Zhang (Huawei); Hongxia Ma (Huawei); Li Zhang (Huawei); Hao Yang (Huawei); Yanfei Jiang (Huawei)

Tutorial on Large Language Models: Principles and Practice [In Theater 1]

Abstract:
The last few years have been marked by several breakthroughs in the domain of generative AI. Large language models such as GPT-4 are able to solve a plethora of tasks, ranging from text and code generation to multimodal data analysis, without task-specific training data.
This tutorial, targeted at database researchers without prior background in language models, introduces language models as well as relevant use cases in the context of data management. The tutorial covers the fundamental principles enabling language models, including the Transformer architecture, pre-training, and alignment. Furthermore, the tutorial will show how to use language models in practice, leveraging OpenAI's GPT model to build a natural language query interface as a demonstration. Finally, the tutorial will discuss recent research exploiting language models in the context of data management.

Presenters:

Immanuel Trummer is an assistant professor at Cornell University and heads the Cornell Database Group. His papers were selected for “Best of VLDB”, “Best of SIGMOD", for the ACM SIGMOD Research Highlight Award, and for publication in CACM as CACM Research Highlight. His online lecture introducing students to database topics collected over a million views. He received the NSF CAREER Award and multiple Google Faculty Research Awards.

Poster Session - [In Speys]

Posters of the 10:30 Session

Demonstrations - Group A - [In B2B]

- Entity/Relationship Profiling
  by Henning Koehler (Massey University), Sebastian Link (University of Auckland)
- GA-Tags: Data Enrichment with an Automatic Tagging System Utilizing Large Language Models
  by Genki Kusano (NEC)
- Comparing Personalized Relevance Algorithms for Directed Graphs
  by Luca Cavalcanti (University of Trento), Cristian Consonni (Joint Research Centre, European Commission), Martin MB Brugnara (University of Trento), David Laniado (Eurecat), Alberto Montresor (University of Trento)
- FSM-Explorer: An Interactive Tool for Frequent Subgraph Pattern Mining from a Big Graph
  by Jalal Khalil (St. Cloud State University), Da Yan (Indiana University Bloomington), Lyuheng Yuan (Indiana University Bloomington), Jiao Han (George Washington University), Saugat Adhikari (Indiana University Bloomington), Cheng Long (Nanyang Technological University), Yang Zhou (Auburn University)
- TASKS: A Real-Time Query System for Instant Error-Tolerant Spatial Keyword Queries on Road Networks
  by Chengyang Luo (Zhejiang University), lu jin (Zhejiang University), Qing Liu (Zhejiang University), Yunjun Gao (Zhejiang University), Lu Chen (Zhejiang University)
- VASIM: Vertical Autoscaling Simulator Toolkit
  by Anna Pavlenko (Microsoft Gray Systems Lab), Karla Saur (Microsoft), Yiwen Zhu (Microsoft), Brian Kroth (Microsoft), Joyce Cahoon (Microsoft), Jesús Camacho-Rodríguez (Microsoft)
- Demonstration of FeVisQA: Free-Form Question Answering over Data Visualization
  by Yuanfeng Song (The Hong Kong University of Science and Technology), 锦伟 卢 (深圳大学), Xuefang Zhao (WeBank Co., Ltd), Raymond Chi-Wing Wong (Hong Kong University of Science and Technology), Haodi Zhang (Shenzhen University)
- CleanEr: Interactive, Query-Guided Error Mitigation for Data Cleaning Systems
  by Ran Schreiber (Bar-Ilan University), Yael Amsterdamer (Bar-Ilan university)
- Wearables for Health (W4H) Toolkit for Acquisition, Storage, Analysis and Visualization of Data from Various Wearable Devices
  by Arash Hajisafi (University of Southern California), Maria Despoina Siampou (University of Southern California), Jize Bi (University of Southern California), Luciano Nocera (University of Southern California), Cyrus Shahabi (Computer Science Department. University of Southern California)

Demonstrations - Group B - [In B2B]

- Chat2Query: A Zero-Shot Automatic Exploratory Data Analysis System with Large Language Models
  by Jun-Peng Zhu (East China Normal University), Peng Cai (East China Normal University), Boyan Niu (PingCAP), Zheming Ni (PingCAP), Kai Xu (PingCAP), Jiajun Huang (PingCAP), Jianwei Wan (PingCAP), Shengbo Ma (PingCAP), Bing Wang (PingCAP), Donghui Zhang (PingCAP), Liu Tang (PingCAP), Qi Liu (PingCAP)
- EADS: An Early Anomaly Detection System for Sensor-based Multivariate Time Series
  by Yihao Ang (National University of Singapore), Qiang Huang (National University of Singapore), Anthony K. H. Tung (NUS), Zhiyong Huang (NUS School of Computing)
- d_{symb} playground: an interactive tool to explore large multivariate time series datasets
  by Sylvain W Combettes (Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli), Paul Boniol (Université de Paris), Charles Truong (Université Paris-Saclay, ENS Paris-Saclay), Laurent Oudre (ENS Paris-Saclay)
- ADecimo: Model Selection for Time Series Anomaly Detection
  by Paul Boniol (Université de Paris), Emmanouil Sylligardos (FORTH), John Paparrizos (The Ohio State University), Panos Trahanias (FORTH), Themis Palpanas (Université Paris Cité)
- ChatGraph: Chat with Your Graphs
  by Yun PENG (Guangzhou University), Sen Lin (Guangzhou University), Qian Chen (Hong Kong Baptist University), Lyu Xu (Hong Kong Baptist University), Xiaojun Ren (Guangzhou University), Yafei Li (Zhengzhou University), Jianliang Xu (Hong Kong Baptist University)
- RLQDAG: a Fast Plan Enumerator for Top-down Recursive Query Optimizers
  by Amela Fejza (Inria), Pierre Genevès (CNRS), Nabil Layaïda (Inria)
- KGSEC: A Modular Framework for Knowledge Graph Schema Extraction and Comparison
  by Petros Skoufis (Athena Research Center), Dimitrios Skoutas (Athena Research Center)
- QFusor: A UDF Optimizer Plugin for SQL Databases
  by Konstantinos Chasialis (Athena RC), Theoni Palaiologou (Athena Research Center), Yannis E Foufoulas (University of Athens), Alkis Simitsis (Athena Research Center), Yannis Ioannidis (University of Athens)
- ARTS: A System for Aggregate Related Table Search
  by Junjie Xing (University of Michigan), H. V. Jagadish (University of Michigan)

Poster Session - [In B2B]

Posters of the PhD Symposium Session

Tuesday May 14th, 2024 18:00-20:00

TCDE Reception - [In Speys]

The Technical Committee on Data Engineering (TCDE), together with the organizing committee of ICDE 2024, is cordialy inviting you for drinks and some snaks and plenty of socializing in the Speys area, right after the last session. At the very end of the Speys area, you will find the TKDE posters. Do not miss the opportunity to talk to their authors who will be waiting for you.

Poster Session - [In Speys]

TKDE Posters

Poster Session - [In B2B]

TKDE Posters

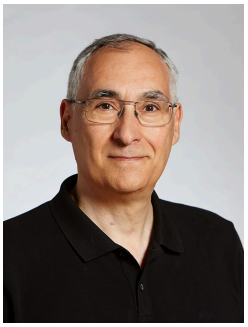
Wednesday May 15th, 2024 8:30-9:30

Keynote 2 [In Theater 12, Chair: TBD]

How serious are we about green computing? The impact of data intensive computing

by Gustavo Alonso (Professor, ETH Zurich).

Many applications dominating the computing landscape are data intensive: data analytics, machine learning, large language models, recommendation systems, etc. The amount of data processed by these systems is staggering and continues to grow at an exponential rate. While the use of more and more data has led to impressive progress in many areas, it has an often ignored side effect: data movement is expensive, requires many resources, and it is often inefficiently managed. Any serious attempt at improving the sustainability and overall efficiency of data centers must necessarily include improvements in the way we handle and process data. In this talk I will show why existing systems are inherently inefficient in data movement, resource utilization, and processing requirements. I will then discuss potential solutions that take advantage of the trends towards specialization and the large economies of scale of the cloud, suggesting along the way how to design data centric architectures that are more energy and resource efficient than what we have today.



Gustavo Alonso is a professor in the Department of Computer Science of ETH Zurich where he is a member of the Systems Group (www.systems.ethz.ch) and the head of the Institute of Computing Platforms. He leads the AMD HACC (Heterogeneous Accelerated Compute Cluster) deployment at ETH (https://github.com/fpgasystems/hacc), with several hundred users worldwide, a research facility that supports exploring data center hardware-software co-design. His research interests include data management, cloud computing architecture, and building systems on modern hardware. Gustavo holds degrees in telecommunication from the Madrid Technical University and a MS and PhD in Computer Science from UC Santa Barbara. Previous to joining ETH, he was a research scientist at IBM Almaden in San Jose, California. Gustavo has received 4 Test-of-Time Awards for his research in databases, software runtimes, middleware, and mobile computing. He is an ACM Fellow, an IEEE Fellow, a Distinguished Alumnus of the Department of Computer Science of UC Santa Barbara, and has received the Lifetime Achievements Award from the European Chapter of ACM SIGOPS (EuroSys).

Founded in 2012, ByteDance' s mission is to inspire creativity and enrich life. With a suite of more than a dozen products, including TikTok, Lark, CapCut, and Lemon8. ByteDance has made it easier and more fun for people to connect with, consume, and create content.

Database Development Team

Our Database Development Team, which falls under the Technical Infrastructure organization, is responsible for building and owning database products including but not limited to relational databases, graph databases, key-value stores and document stores, which are used across ByteDance. In this team, you'll have the opportunity to develop and enhance a full-stack database management system including storage, query execution engine and query optimizer as well as associated management tool in a cloud native environment. We embrace a culture of self- direction, intellectual curiosity, openness, and problem-solving.

Infrastructure System Lab

The Infrastructure System Lab in the Technical Infrastructure team works on cutting-edge infrastructure system innovations. Along with the development teams, the lab leads state-of-the-art technology breakthroughs in fields that then lead to the creation of powerful backbones for ByteDance products and corporate partners. These fields includebut are not limited to computing,storage, database, network, and security.

Data Architecture Team

Data Platform-Data Architecture team of ByteDance focuses on building the next generation DBMS kernel aiming for the best-in-class query latency and hardware utilization. The team operates the distributed OLAP infrastructure powering cloud native data warehouse, which is the foundation behind data-driven decision making within ByteDance.

Let's Stay Connected | @ByteDancer @ByteDance @ByteDance

Wednesday May 15th, 2024 9:30-10:00

Coffee Break - [In Speys]

Wednesday May 15th, 2024 10:00-12:15

Database technology for AI I - [In Theater 12, Chair: Jia Zou]

- SparDL: Distributed Deep Learning Training with Efficient Sparse Communication
  Minjun Zhao (Zhejiang University); Yichen Yin (Zhejiang University); Yuren Mao (Zhejiang University); Qing Liu (Zhejiang University); Lu Chen (Zhejiang University); Yunjun Gao (Zhejiang University)
- MetaSQL: A Generate-then-Rank Framework for Natural Language to SQL Translation
  Yuankai Fan (Fudan University); Zhenying He (Fudan University); Tonghui Ren (Fudan University); Can Huang (Fudan University); Yinan Jing (Fudan University); Kai Zhang (Fudan University); X. Sean Wang (Fudan University)
- Feed: Towards Personalization-Effective Federated Learning
  Pengpeng Qiao (Beijing Institute of Technology); Kangfei Zhao (Beijing Insitute of Technology); Bei Bi (Beijing Institute of Technology); Zhiwei Zhang (Beijing Institute of Technology); Ye Yuan (Beijing Institute of Technology); Guoren Wang (Beijing Institute of Technology)
- T-Rex (Tree-Rectangles): Reformulating Decision Tree Traversal as Hyperrectangle Enclosure
  Meghana Madhyastha (Johns Hopkins University); Tamas Budavari (Johns Hopkins University); Joshua T Vogelstein (Johns Hopkins); Vladmir Braverman (Rice University); Randal Burns (Johns Hopkins University)
- FeatAug: Automatic Feature Augmentation From One-to-Many Relationship Tables
  Danrui Qi (Simon Fraser University); Weiling Zheng (Simon Fraser University); Jiannan Wang (Simon Fraser University)
- AutoMC: Automated Model Compression Based on Domain Knowledge and Progressive Search
  Chunnan Wang (Tencent Inc.); Hongzhi Wang (Harbin Institute of Technology); Xiangyu Shi (Harbin Institute of Technology)
- Task-Oriented GNNs Training on Large Knowledge Graphs for Accurate and Efficient Modeling
  Hussein Shahata Abdallah (Concordia University); Waleed Afandi (Concordia University); Panos Kalnis (King Abdullah University of Science and Technology); Essam Mansour (Concordia University)

Distributed, Parallel and P2P Data Management I - [In Theater 11, Chair: Yongluan Zhou]

- Lion: Minimizing Distributed Transactions through Adaptive Replica Provision
  Qiushi Zheng (Renmin University of China); Zhanhao Zhao (Renmin University of China); WEI LU (Renmin University of China); Chang Yao (Zhejiang University); Yuxing Chen (Tencent); Anqun Pan (Tencent Inc., China); Xiaoyong Du (Renmin University of China)
- FC: Adaptive Atomic Commit via Failure Detection
  Hexiang Pan (National University of Singapore); Quang-Trung Ta (National University of Singapore); Meihui Zhang (Beijing Institute of Technology); Zhanhao Zhao (National University of Singapore); Yeow Meng Chee (National University of Singapore); Gang Chen (Zhejiang University); Beng Chin Ooi (NUS)
- ZeroTune: Learned Zero-Shot Cost Model for Parallelism Tuning in Stream Processing
  Pratyush Agnihotri (TU Darmstadt); Boris Koldehofe (TU Ilmenau); Paul Stiegele (TU Darmstadt); Roman Heinrich (DHBW Mannheim); Carsten Binnig (TU Darmstadt); Manisha Luthra (TU Darmstadt and DFKI)
- MergeSFL: Split Federated Learning with Feature Merging and Batch Size Regulation
  Yunming Liao (University of Science and Technology of China); Yang Xu (University of Science and Technology of China); Hongli Xu (University of Science and Technology of China); Lun Wang (University of Science and Technology Of China); zhiwei yao (University of Science and Technology of China); Chunming Qiao (University at Buffalo)
- SharDAG: Scaling DAG-Based Blockchains via Adaptive Sharding
  Feng Cheng (Huazhong University of Science and Technology); Jiang Xiao (Huazhong University of Science and Technology); liu cunyang (Huazhong University of Science and Technology); Shijie Zhang (Huazhong University of Science and Technology); Yifan Zhou (Huazhong University of Science and Technology); Bo Li (Hong Kong University of Science and Technology); Baochun Li (University of Toronto); Hai Jin (Huazhong University of Science and Technology)
- Boosting Write Performance of KV Stores: An NVM-Enabled Storage Collaboration Approach
  Yi Wang (Shenzhen University); Jiajian He (Shenzhen University); Kaoyi Sun (Shenzhen University); Yunhao Dong (Shenzhen University); Jiaxian Chen (Shenzhen University); Chenlin Ma (Shenzhen University); Amelie Chi Zhou (Hong Kong Baptist University); Rui Mao (Shenzhen University)
- Log Replaying for Real-Time HTAP: An Adaptive Epoch-based Two-Stage Framework
  Jun-Peng Zhu (East China Normal University); Zhiwei Ye (China Mobile Communications Corporation); Peng Cai (East China Normal University); Donghui Wang (East China Normal University); Zhang Fengyan (ECUN); Dunbo Cai (hina Mobile (Suzhou) Software Technology Co., Ltd.); Ling Qian (China Mobile Cloud Computing Center)

Explainability, Fairness, and Trust in Data Systems I [In Theater 10, Chair: Dim. Sacharidis]

- Mitigating Subgroup Unfairness in Machine Learning Classifiers: A Data-Driven Approach
  Yin Lin (University of Michigan); Samika Gupta (University of Michigan); H. V. Jagadish (University of Michigan)
- Non-Invasive Fairness in Learning through the Lens of Data Drift
  Ke Yang (University of Texas at San Antonio); Alexandra Meliou (University of Massachusetts Amherst)
- Preventing the Popular Item Embedding Based Attack in Federated Recommendations
  Jun Zhang (Zhejiang University); Huan Li (Zhejiang University); Dazhong Rong (Zhejiang University); Yan Zhao (Aalborg University); Ke Chen (Zhejiang University); Lidan Shou (Zhejiang University)
- Explainable Disparity Compensation for Efficient Fair Ranking
  Abraham Gale (Rutgers University); Amelie Marian (Rutgers University)
- Generating Explanations to Understand and Repair Embedding-based Entity Alignment
  Xiaobin Tian (Nanjing University); Zequn Sun (Nanjing University); Wei Hu (Nanjing University)
- Inspection Before Instruction: Self-Rationalization with a Discriminative Cooperative Game
  Wei Liu (Huazhong University of Science and Technology); Haozhao Wang (Huazhong University of Science and Technology); Jun Wang (iWudao); Zhiying Deng (Huazhong University of Science and Technology); YuanKai Zhang (Huazhong University of Science and Technology); Cheng Wang (Huazhong University of Science and Technology); Ruixuan Li (Huazhong University of Science and Technology)
- Model Trip: Enhancing Privacy and Fairness in Model Fusion across Multi-Federations for Trustworthy Global Healthcare
  Qian Chen (Institute of Computing Technology, Chinese Academy of Sciences); Yiqiang Chen (Institute of Computing Technology, Chinese Academy of Sciences); Bingjie Yan (Institute of Computing Technology, Chinese Academy of Sciences); Xinlong JIANG (Institute of Computing Technology, Chinese Academy of Sciences); Xiaojin Zhang (HUST); Yan Kang (Webank); Teng Zhang (Institute of Computing Technology, CAS); Wuliang Huang (ICT,CAS); Chenlong Gao (Institute of Computing Technology, Chinese Academy of Sciences); Lixin Fan (WeBank); Qiang Yang (Hong Kong UST)

Data Mining and Knowledge Discovery VI - [In Theater 9, Chair: Tsz Nam Chan]

- Learning Multi-Pattern Normalities in the Frequency Domain for Efficient Anomaly Detection
  Feiyi Chen (Zhejiang University); Yingying Zhang (Alibaba Group); Zhen Qin (Zhejiang University); Lunting Fan (Alibaba Group); Renhe Jiang (The University of Tokyo); Yuxuan Liang (The Hong Kong University of Science and Technology (Guangzhou)); Qingsong Wen (Squirrel AI); Shuiguang Deng (Zhejiang University)
- Modeling User Attention in Music Recommendation
  Sunhao Dai (Renmin University of China); Ninglu Shao (Renmin University of China); Jieming Zhu (Huawei Noah's Ark Lab); Xiao Zhang (Renmin University of China); Zhenhua Dong (Huawei Noah's Ark Lab); Jun Xu (Renmin University of China); Quanyu Dai (Huawei Noah's Ark Lab); Ji-Rong Wen (Renmin University of China)
- A Robust Prioritized Anomaly Detection when Not All Anomalies are of Primary Interest
  Guanyu Lu (East China Normal University); Fang Zhou (East China Normal University); Martin Pavlovski (Yahoo Inc.); Chenyi Zhou (East China Normal University); Cheqing Jin (East China Normal University)
- Enhancing Quantitative Reasoning Skills of Large Language Models through Dimension Perception
  Yuncheng Huang (Fudan University); Qianyu He (Fudan University); Jiaqing Liang (Fudan University); Sihang Jiang (Fudan University); Yanghua Xiao (Fudan University); Yunwen Chen (DataGrand Inc.)
- SSDRec: Self-Augmented Sequence Denoising for Sequential Recommendation
  Chi Zhang (Harbin Engineering University); Qilong Han (Harbin Engineering University); Rui Chen (Harbin Engineering University); Xiangyu Zhao (City University of Hong Kong); Peng Tang (Shandong University); Hongtao Song (Harbin Engineering University)
- BSL: Understanding and Improving Softmax Loss for Recommendation
  Junkang Wu (University of Science and Technology of China); Jiawei Chen (Zhejiang University); Jiancan Wu (University of Science and Technology of China); Wentao Shi (University of technology and science of China); Jizhi Zhang (University of Science and Technology of China); Xiang Wang (National University of Singapore)
- Online Detection of Outstanding Quantiles with QuantileFilter
  Yuhan Wu (Peking University); Aomufei Yuan (Peking University); Zhouran Shi (Peking University); Yuanpeng Li (Peking University); Yikai Zhao (Peking University); Peiqing Chen (Peking University); Tong Yang (Peking University); Bin Cui (Peking University)

Crowdsourcing I - [In Theater 4, Chair: Andrew Rebeiro-Hargrave]

- HITSNDIFFS: From Truth Discovery to Ability Discovery by Recovering Matrices with the Consecutive Ones Property
  Zixuan Chen (Northeastern University); Subhodeep Mitra (Google); R Ravi (CMU); Wolfgang Gatterbauer (Northeastern University)
- Cross-domain-aware Worker Selection with Training for Crowdsourced Annotation
  Yushi Sun (Hong Kong University of Science and Technology); Jiachuan Wang (HKUST); Peng Cheng (East China Normal University); Libin Zheng (Sun Yat-sen University); Lei Chen (Hong Kong University of Science and Technology); Jian Yin (Sun Yat-Sen University)
- Graph Contrastive Learning for Truth Inference
  Hao Liu (East China Normal Univeisity); Jiacheng Liu (Chinese University of Hong Kong); Feilong Tang (Shanghai Jiao Tong University); Peng Li (the University of Aizu); Long Chen (Shanghai Jiao Tong University); Jiadi Yu (Shanghai Jiao Tong University); Yanmin Zhu (Shanghai Jiao Tong University); Yanqin Yang (East China Normal Univeisity); Ming Gao (East China Normal University); Xiaofeng Hou (Shanghai Jiao Tong University)
- Task Recommendation in Spatial Crowdsourcing: A Trade-off between Diversity and Coverage
  Liwei Deng (University of Electronic Science and Technology of China); Yan Zhao (Aalborg University); Yue Cui (The Hong Kong University of Science and Technology); Yuyang Xia (University of Electronic Science and Technology of China); Jin Chen (University of Electronic Science and Technology of China(UESTC)); Kai Zheng (University of Electronic Science and Technology of China)
- MACRO: Incentivizing Multi-leader Game-based Pareto-efficient Crowdsourcing for Video Analytics
  Yu Chen (Nanjing University); Sheng Zhang (Nanjing University); Ziying Zhou (Beijing Normal University); Xiaokun Wang (Nanjing University); Yu Liang (Nanjing Normal University); Ning Chen (Nanjing University); Yuting Yan (Nanjing University); Mingjun Xiao (University of Science and Technology of China); Jie Wu (Temple University); Zhuzhong Qian (NJU); Harry Xu (UCLA)
- Cooperative Global Path Planning for Multiple Platforms
  Xiaoxi Cui (Beijing Institute of Technology); Yurong Cheng (Beijing institute of technology); Siyi Zhang (Beijing Institute of Technology); Ye Yuan (Beijing Institute of Technology); Guoren Wang (Beijing Institute of Technology)
- Cross Online Assignment of Hybrid Task in Spatial Crowdsourcing
  Zhao Liu (Hunan university); Guoqing Xiao (Hunan University); Xu Zhou (Hunan university); qin yunchuan (hunan university); Yunjun Gao (Zhejiang University); Kenli Li (Hunan University)

Diversity, Equality and Inclusion in DB Venues [In Theater 3, Chair: Pinar Tözün, Katerina Ioannou]

- Diversity, Equality and Inclusion in Software Engineering
  by Prof. Alexander Serebrenik (Eindhoven University of Technology)

HUAWEI

Huawei Cloud GaussDB: A Better Way to Database

High security

CC EAL4+
Industry's highest certification

Fully encrypted database

Anti-tamper database

High availability

Dual-cluster strong consistency

Strong consistency between clusters

Application Lossless and Transparent

High scalability

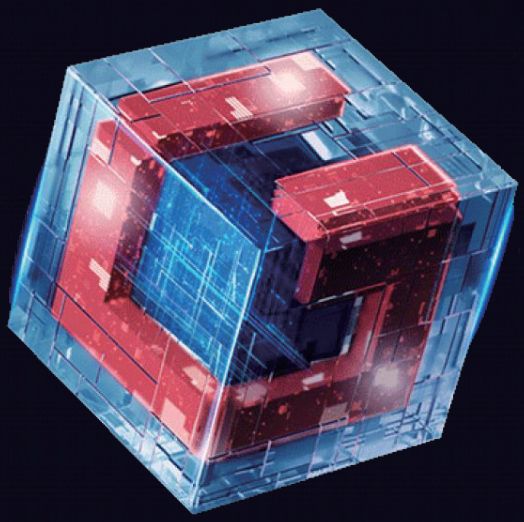
Up to 1,000+ nodes per distributed cluster

Online cluster scale-out

High performance

An innovative industry-leading storage engine

Ustore storage engine



High intelligence

AI-native database

SQL Audit

Intelligent O&M

Smart deployment

Lightweight deployment | Data compression | Multi-tenancy

Smart migration

Compatible with common Oracle/MySQL syntax

Industry - Infrastructure for Machine Learning - [In Theater 2, Chair: Peter Boncz]

- Invited Talk: Deletion Vectors: No-Regrets Row-Level Updates in Delta Lake
  Bart Samwel (Databricks)
- GaussML: An End-to-End In-database Machine Learning System
  Guoliang Li (Tsinghua University); Ji Sun (Huawei); Lijie Xu (ETH Zurich); Shifu Li (Huawei); Jiang Wang (Huawei); Wen Nie (Huawei)
- Xorbits: Automating Operator Tiling for Distributed Data Science
  Weizheng Lu (Renmin University of China); Kaisheng He (Xorbits Inc.); Xuye Qin (Xorbits Inc.); Chengjie Li (Xorbits Inc.); Zhong Wang (Xorbits); Tao Yuan (China Communications Information Technology Group Co. , Ltd); Feng Zhang (Renmin University of China); Yueguo Chen (Renmin University of China); Xiaoyong Du (Renmin University of China)
- Couler: Unified Machine Learning Workflow Optimization in Cloud
  Xiaoda Wang (Sichuan University); Yuan Tang (Uptake Technologies); Tengda Guo (Sichuan University); Jian Sha (AntGroup); Bo Sang (AntGroup); jingji wu (AntGroup); Ke Zhang (Ant Group); Jiang Qian (Snap); Mingjie Tang (Sichuan University)*
- AntDT: A Self-Adaptive Distributed Training Framework for Leader and Straggler Nodes
  Youshao Xiao (Ant Group); Lin Ju (Ant Group); Zhenglei Zhou (Ant Group); Siyuan Li (Ant Group); Zhaoxin Huan (Ant Financial Services Group); Dalong Zhang (Ant Group); Rujie Jiang (Ant Group); Lin Wang (Ant Group); Xiaolu Zhang (Ant Financial Services Group); Lei Liang (Ant Group); Jun Zhou (Ant Financial)

Tutorial on Privacy-Aware Analysis based on Data Series [In Theater 1]

Abstract:
Data that is recorded about the operations of an organization constitutes a valuable source of information for monitoring and improvement. Specific use cases include the assessment of compliance to legal regulations, the analysis of performance bottlenecks, or the optimization of resource utilization. In recent years, a plethora of algorithms for operational analysis using data series, summarized as process mining, have been developed to support these use cases. Data series often contain sensitive information, though, about the individuals that act as service consumers or service providers. Personal information is only partially hidden by obfuscation and pseudonymization and potential privacy breaches need to be prevented for ethical, legal, and economic reasons. This tutorial is devoted to methods for privacy-aware analysis using data series. It covers essential notions, reviews privacy-disclosure attacks, and outlines techniques to give formal privacy guarantees while largely maintaining the data's utility for operational analysis. The discussion is structured by the adopted perspective on the privacy of individuals, and the degree to which a data series contains contextual information.

Presenters:

Stephan Fahrenkrog-Petersen is a research group lead at the Weizenbaum Institute, Germany. He holds a PhD from Humboldt-Universität zu Berlin. His research was published in the proceedings of the premier conferences in the field and in international journals, such as ACM TMIS, DKE, and KAIS. His work received the Distinguished Paper Award at CAiSE 2021 and the Best Student Paper Award at ICPM 2021.

Han van der Aa is a junior professor in the Data and Web Science Group at the University of Mannheim, Germany. He obtained a PhD from the Vrije Universiteit Amsterdam in 2018. His research interests include process modelling, process mining, natural language processing, and complex event processing. His work has been published in journals including IEEE TKDE, Information Systems, and Decision Support Systems and at the BPM, CAISE, ICPM, ICDE, and SIGMOD conferences.

Matthias Weidlich is a professor and Chair of Databases and Information Systems at Humboldt-Universität zu Berlin, Germany. Matthias' research focuses on process-oriented and event-based information systems. His results appear regularly in premier conferences (SIGMOD, VLDB, ICDE, IJCAI, AAAI, BPM, CAiSE) and journals (TKDE, Information Systems, VLDB Journal) in the field. He serves as Co-Editor-in-Chief for the Information Systems journal and is a member of the steering committee of the ACM DEBS conference series.

Poster Session - [In Speys]

Posters of the 15:30 Session

Demonstrations - Group B - [In B2B]

- Chat2Query: A Zero-Shot Automatic Exploratory Data Analysis System with Large Language Models
  by Jun-Peng Zhu (East China Normal University), Peng Cai (East China Normal University), Boyan Niu (PingCAP), Zheming Ni (PingCAP), Kai Xu (PingCAP), Jiajun Huang (PingCAP), Jianwei Wan (PingCAP), Shengbo Ma (PingCAP), Bing Wang (PingCAP), Donghui Zhang (PingCAP), Liu Tang (PingCAP), Qi Liu (PingCAP)
- EADS: An Early Anomaly Detection System for Sensor-based Multivariate Time Series
  by Yihao Ang (National University of Singapore), Qiang Huang (National University of Singapore), Anthony K. H. Tung (NUS), Zhiyong Huang (NUS School of Computing)
- d_{symb} playground: an interactive tool to explore large multivariate time series datasets
  by Sylvain W Combettes (Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli), Paul Boniol (Université de Paris), Charles Truong (Université Paris-Saclay, ENS Paris-Saclay), Laurent Oudre (ENS Paris-Saclay)
- ADecimo: Model Selection for Time Series Anomaly Detection
  by Paul Boniol (Université de Paris), Emmanouil Sylligardos (FORTH), John Paparrizos (The Ohio State University), Panos Trahanias (FORTH), Themis Palpanas (Université Paris Cité)
- ChatGraph: Chat with Your Graphs
  by Yun PENG (Guangzhou University), Sen Lin (Guangzhou University), Qian Chen (Hong Kong Baptist University), Lyu Xu (Hong Kong Baptist University), Xiaojun Ren (Guangzhou University), Yafei Li (Zhengzhou University), Jianliang Xu (Hong Kong Baptist University)
- RLQDAG: a Fast Plan Enumerator for Top-down Recursive Query Optimizers
  by Amela Fejza (Inria), Pierre Genevès (CNRS), Nabil Layaïda (Inria)
- KGSEC: A Modular Framework for Knowledge Graph Schema Extraction and Comparison
  by Petros Skoufis (Athena Research Center), Dimitrios Skoutas (Athena Research Center)
- QFusor: A UDF Optimizer Plugin for SQL Databases
  by Konstantinos Chasialis (Athena RC), Theoni Palaiologou (Athena Research Center), Yannis E Foufoulas (University of Athens), Alkis Simitsis (Athena Research Center), Yannis Ioannidis (University of Athens)
- ARTS: A System for Aggregate Related Table Search
  by Junjie Xing (University of Michigan), H. V. Jagadish (University of Michigan)

Demonstrations - Group C - [In B2B]

- Explaining Expert Search Systems with ExES
  by Kiarash Golzadeh (University of Waterloo), Lukasz Golab (University of Waterloo), Jaroslaw Szlichta (York University and IBM CAS)
- RAGE Against the Machine: Retrieval-Augmented LLM Explanations
  by Joel E Rorseth (University of Waterloo), Parke Godfrey (York University), Lukasz Golab (University of Waterloo), Divesh Srivastava (AT&T Chief Data Office), Jaroslaw Szlichta (York University and IBM CAS)
- FairCR - an evaluation and recommendation system for fair classification algorithms
  by Nico Lässig (University of Stuttgart), Melanie Herschel (Universität Stuttgart)
- GraphLingo: Domain Knowledge Exploration by Synchronizing Knowledge Graphs and Large Language Models
  by Duy Le (Case Western Reserve University), Kris Zhao (Case Western Reserve University), Mengying Wang (Case Western Reserve University), Yinghui Wu (Case Western Reserve University)
- MixedSearch: An Interactive System of Searching for the Best Tuple with Mixed Attributes
  by Weicheng Wang (Hong Kong University of Science and Technology), Min Xie (Shenzhen Institute of Computing Sciences), Raymond Chi-Wing Wong (Hong Kong University of Science and Technology)
- MorphStream: Scalable Processing of Transactions over Streams
  by Siqi Xiang (Singapore University of Technology and Design), Zhonghao Yang (Singapore University of Technology and Design), Shuhao Zhang (Nanyang Technological University), Jianjun Zhao (Huazhong University of Science and Technology), Yancan Mao (National University of Singapore)
- FONT: A Flexible Polystore Evaluation Platform
  by Gengyuan Shi (Tsinghua University), Chaokun Wang (Tsinghua University), Minghao Zhang (Tsinghua University), Binbin Wang (MSFT)
- CAMO: Explaining Consensus Across MOdels
  by Andy Yu (Waterloo University), Parke Godfrey (York University), Lukasz Golab (University of Waterloo), Divesh Srivastava (AT&T Chief Data Office), Jaroslaw Szlichta (York University and IBM CAS)
- Pyneapple-R: Scalable and Expressive Spatial Regionalization
  by Yunfan Kang (University of California Riverside), Yongyi Liu (University of California, Riverside), Hussah Alrashid (University of California Riverside), Akash Bilgi (University of California, Riverside), Siddhant Purohit (University of California, Riverside), Ahmed Mahmood (Google), Sergio Rey (San Diego State University), Amr Magdy (University of California Riverside)

Poster Session - [In B2B]

Posters of the PhD Symposium Session

Wednesday May 15th, 2024 12:15-13:45

Lunch - [In Speys]

Wednesday May 15th, 2024 13:45-15:15

Executive and Awards Session [In Theater 12, Chair: TBD]

Best Paper Award
Reverse Regret Query
by Weicheng Wang, Raymond Chi-Wing Wong, H. V. Jagadish, and Min Xie

Best Runner Up
LBSC: A Cost-aware Caching Framework for Cloud Databases
by Zhaoxuan Ji, Zhongle Xie, Yuncheng Wu, Meihui Zhang

Best Industry and Application Paper Award
Towards a Shared-storage-based Serverless Database Achieving Seamless Scale-up and Read Scale-out
by Yingqiang Zhang, Xinjun Yang, Hao Chen, Feifei Li, Jiawei Xu, Jie Zhou, Xudong Wu, and Qiang Zhang

Best Industry and Application Paper Runner Up
Xorbits: Automating Operator Tiling for Distributed Data Science
by Weizheng Lu, Kaisheng He, Xuye Qin, Chengjie Li, Zhong Wang, Tao Yuan, Feng Zhang, Yueguo Chen, and Xiaoyong Du

IEEE TCDE Education Award
Tamer Özsu, Univ of Waterloo
For fundamental contributions to data management and data science pedagogy.

IEEE TCDE Rising Star Award
Natacha Crooks, UC Berkeley
For contributions to distributed data management, and its applications to blockchain technology, security, and cloud computing.

IEEE TCDE Impact Award
Sihem Amer-Yahia, CNRS, Grenoble
For contributions to laying the interdisciplinary data engineering foundations of social computing and crowdsourcing.

IEEE T&C Distinguished Service Award
Malu Castellanos, Teradata USA
for her essential contributions both as a TCDE Executive Committee member and a member of the ICDE Steering Committee.

IEEE Innovation in Societal Infrastructure Award
Elena Ferrari, University of Insubria
For pioneering and sustained contributions to the security and privacy of online social networks.

OCEANBASE

Unlimited scalable distributed database for data-intensive transactional and real-time operational analytics workloads, with ultra-fast performance that has once achieved world records in the TPC-C and TPC-H benchmark test. OceanBase has served over 1000 customers across the globe and has been supporting all mission critical systems in Alipay.

 # 1

The only distributed database that once topped the TPC-C and TPC-H benchmark tests.

610 M

Supporting Double Eleven Shopping Festival for 11 years, with 610 million QPS at peak time.

1000+

Serving over 1000 customers and partners from various industries across the globe.

Built for Mission-Critical Workloads

Ultra Scalable

Support on-demand elastic scaling, achieving transparent scaling without application changes or downtime, effortlessly handling traffic spikes. Support multi-read and multi-write capabilities.

High Availability

Achieve RPO = 0 and RTO < 8 seconds, enabling automatic disaster recovery through a distributed multi-replica architecture during failure of a server, rack or region, effectively eliminating downtime.

Real-time Analytics

Empower a single database and data set to seamlessly handle both transactional and real-time analytical workloads using enterprise-grade optimizer for complex SQL optimization and processing.

Cloud Agnostic & Seamless Migration

Support multi-cloud deployments on various CSPs including AWS, GCP and Alibaba Cloud. Highly compatible with MySQL as well as other leading commercial databases, ensuring seamless migration with minimal modifications.

Cost-effectiveness at Scale

Reducing storage costs by 70%-90% with advanced compression technology while improving resource utilization with multi-tenancy capabilities, OceanBase slashes TCO by over 30% in high-concurrency scenarios.

Our Publications

Our team are dedicated to delivering impactful and groundbreaking insights to everyone by sharing our research and tools to fuel progress in the database and the distributed technologies.

IEEE ICDE
International Conference on Data Engineering

- Functionality-Aware Database Tuning via Multi-Task Learning (ICDE 2024)
- BenchTemp: A General Benchmark for Evaluating Temporal Graph Neural Networks (ICDE 2024)
- Generative and Contrastive Paradigms Are Complementary for Graph Self-Supervised Learning (ICDE 2024)
- LCL: A Lock Chain Length-based Distributed Algorithm for Deadlock Detection and Resolution (ICDE 2023)

ACM SIGMOD

- F3KM: Federated, Fair, and Fast k-means (SIGMOD 2024)
- FEAST: A Communication-efficient Federated Feature Selection Framework for Relational Data (SIGMOD 2023)
- Dike: A Benchmark Suite for Distributed Transactional Databases (SIGMOD 2023)

Very Large Data Bases

- OceanBase Paetica: A Hybrid Shared-nothing / Shared-everything Database for Supporting Single Machine and Distributed Cluster (VLDB 2023)
- Efficient Distributed Transaction Processing in Heterogeneous Networks (VLDB 2023)
- OceanBase: A 707 Million tpmC Distributed Relational Database System (VLDB 2022)

Delving into Database Frontiers with OceanBase Lab

OceanBase Lab is dedicated to exploring the frontier core technologies of databases, expanding the boundaries of database technology, and offering customers a reservoir of technical strength for superior database products. The Lab investigates industry-leading distributed HTAP database technology, addressing practical database challenges.

This effort significantly enhances OceanBase's influence within the database industry. With a diverse portfolio of research projects, we are eager to communicate and collaborate with premier research institutions in the database field.

Join our Team: **https://bit.ly/3WdZP88**
For further inquiries and research opportunities, please contact **xuquanqing.xqq@oceanbase.com**

Wednesday May 15th, 2024 15:30-16:30

Database technology for AI II - [In Theater 12, Chair: Long Yuan]

- Clients Help Clients: Alternating Collaboration for Semi-Supervised Federated Learning
  Zhida Jiang (University of Science and Technology of China); Yang Xu (University of Science and Technology of China); Hongli Xu (University of Science and Technology of China); Zhiyuan Wang (University of Science and Technology of China); Chunming Qiao (University at Buffalo)
- AutoFeat: Transitive Feature Discovery over Join Paths
  Andra Ionescu (TU Delft); Kiril Vasilev (TU Delft); Florena Buse (TU Delft); Rihan Hai (TU Delft); Asterios Katsifodimos (TU Delft)
- Triple-d: Denoising Distant Supervision for High-quality Data Creation
  Xinyi ZHU (HKUST (GZ)); Yongqi Zhang (4Paradigm Inc.); Lei Chen (Hong Kong University of Science and Technology); Kai Chen (HKUST)

Distributed, Parallel and P2P Data Management II - [In Theater 11, Chair: Nikos Ntarmos]

- FSD-Inference: Fully Serverless Distributed Inference with Scalable Cloud Communication
  Joe Oakley (University of Warwick); Hakan Ferhatosmanoglu (University of Warwick)
- Graph Computation with Adaptive Granularity
  Ruiqi Xu (National University of Singapore); Yue Wang (Shenzhen Institute of Computing Sciences); Xiaokui Xiao (National University of Singapore)
- FedCross: Towards Accurate Federated Learning via Multi-Model Cross-Aggregation
  Ming Hu (Nanyang Technological University); pei heng zhou (ECNU); Zhihao Yue (East China Normal University); zhiwei ling (East China Normal University); Yihao Huang (Nanyang Technological University); Anran Li (Nanyang Technological University); Yang Liu (Nanyang Technology University, Singapore); Xiang Lian (Kent State University); Mingsong Chen (East China Normal University)

Explainability, Fairness, and Trust in Data Systems and Analysis II - [In Theater 10, Chair: Paolo Missier]

- Why-Not Explainable Graph Recommender
  Hervé-Madelein Attolou (ETIS - CY); Katerina Tzompanaki (CY Cergy Paris University); Kostas Stefanidis (Tampere University); Dimitris Kotzinos (CY Cergy Paris University)
- GAGE: Genetic Algorithm-based Graph Explainer for Malware Analysis
  Mohd Saqib (McGill University); Benjamin C. M. Fung (McGill University); Philippe Charland (Mission Critical Cyber Security Section, Defence R&D Canada - Valcartier); Andrew Walenstein (BlackBerry Limited)
- Accurate Explanation Model for Image Classifiers using Class Association Embedding
  Ruitao Xie (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, 518055, China); Jingbang Chen (University of California, Los Angeles); Limai Jiang (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences); Rui Xiao (Shenzhen institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen,518055,China); Yi Pan (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences); Yunpeng Cai (Shenzhen Institutes of Advanced Technology)

Data Mining and Knowledge Discovery VII - [In Theater 9, Chair: Luo Zhaojing]

- When Multi-Behavior Meets Multi-Interest: Multi-Behavior Sequential Recommendation with Multi-Interest Self-Supervised Learning
  Binquan Wu (South China University of Technology); Yu Cheng (South China University of Technology); Haitao Yuan (South China University of Technology); Qianli Ma (South China University of Technology)
- E2GCL: Efficient and Expressive Contrastive Learning on Graph Neural Networks
  Haoyang Li (The Hong Kong University of Science and Technology); Shimin Di (The Hong Kong University of Science and Technology); Lei Chen (Hong Kong University of Science and Technology); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- Ambiguous Entity Oriented Targeted Document Detection
  Wei Shen (Nankai University); Haixu Wen (Nankai University)

Crowdsourcing II - [In Theater 4, Chair: Yan Zhao]

- RA3: Human-in-the-loop Framework for Interpreting and Improving Image Captioning with Relation-Aware Attribution Analysis
  Lei Chai (Beihang University); Lu Qi (Beihang University); Hailong Sun (Beihang University, China); Jingzheng Li (Zhongguancun Laboratory)
- Efficient Example-Guided Interactive Graph Search
  Zhuowei Zhao (The University of Melbourne); Junhao Gan (University of Melbourne); Jianzhong Qi (The University of Melbourne); Zhifeng Bao (RMIT University)
- Wait to be Faster: a Smart Pooling Framework for Dynamic Ridesharing
  Xiaoyao Zhong (East China Normal University); Jin Jiabao (East China Normal University); Peng Cheng (East China Normal University); Wangze Ni (Hong Kong University of Science and Technology); Libin Zheng (Sun Yat-sen University); Lei Chen (Hong Kong University of Science and Technology); Xuemin Lin (Shanghai Jiaotong University)

Database technology for Blockchains I - [In Theater 3, Chair: Uta Störl]

- Efficient Partial Order Based Transaction Processing for Permissioned Blockchains
  Shuai Zhao (Beijing Institute of Technology); Zhiwei Zhang (Beijing Institute of Technology); Junkai Wang (Beijing Institute of Technology); Ye Yuan (Beijing Institute of Technology); Meihui Zhang (Beijing Institute of Technology); Guoren Wang (Beijing Institute of Technology); Jiang Xiao (Huazhong University of Science and Technology)
- TELL: Efficient Transaction Execution Protocol Towards Leaderless Consensus
  Xing Tong (East China Normal University); Zheming Ye (East China Normal University); Zhao Zhang (East China Normal University); Cheqing Jin (East China Normal University); Aoying Zhou (East China Normal University)
- SpotLess: Concurrent Rotational Consensus Made Practical through Rapid View Synchronization
  Dakai Kang (University of California, Davis); Sajjad Rahnama (UC Davis); Jelle Hellings (McMaster University); Mohammad Sadoghi (University of California, Davis)
- PrestigeBFT: Revolutionizing View Changes in BFT Consensus Algorithms with Reputation Mechanisms
  Gengrui Zhang (University of Toronto); Fei Pan (University of Toronto); Sofia Tijanic (University of Toronto); Hans-Arno Jacobsen (University of Toronto)

Data Mining and Knowledge Discovery VIII - [In Theater 2, Chair: Wenzhong Li]

- TS3Net: Triple Decomposition with Spectrum Gradient for Long-Term Time Series Analysis
  xiangkai ma (Nanjing University); Xiaobin Hong (Nanjing university); Sanglu Lu (NJU); Wenzhong Li (Nanjing University)
- An Efficient Fuzzy Stream Clustering Method Based on Granular-Ball Structure
  Jiang Xie (Chongqing Key Laboratory of Computational Intelligence; Chongqing University of Posts and Telecommunications); Minggao Dai (Chongqing Key Laboratory of Computational Intelligence; Chongqing University of Posts and Telecommunications); Shuyin Xia Xia (Chongqing University of Posts and Telecommunications); Jingjing Zhang (Chongqing Key Laboratory of Computational Intelligence; Chongqing University of Posts and Telecommunications); Guoyin Wang (Chongqing Key Laboratory of Computational Intelligence; Chongqing University of Posts and Telecommunications); Xinbo Gao (Chongqing University of Posts and Telecommunications)
- W-GBC: An Adaptive Weighted Clustering Method Based on Granular-Ball Structure
  Jiang Xie (Chongqing Key Laboratory of Computational Intelligence; Chongqing University of Posts and Telecommunications); Yuxin Cheng (Chongqing Key Laboratory of Computational Intelligence; Chongqing University of Posts and Telecommunications); Shuyin Xia Xia (Chongqing University of Posts and Telecommunications); Chunfeng Hua (Chongqing Key Laboratory of Computational Intelligence; Chongqing University of Posts and Telecommunications); Guoyin Wang (Chongqing Key Laboratory of Computational Intelligence; Chongqing University of Posts and Telecommunications); Xinbo Gao (Chongqing University of Posts and Telecommunications)

Data Mining and Knowledge Discovery IX - [In Theater 1, Chair: Chuan Peng]

- RobFL: Robust Federated Learning via Feature Center Separation and Malicious Center Detection
  Ting Zhou (Shandong University); Ning Liu (School of Software, Shandong University); Bo Song (Shandong University); Hongtao Lv (Shandong University); Deke Guo (National University of Defense Technology); Lei LIU (Shandong University)
- Towards Task-Conflicts Momentum-Calibrated Approach for Multi-task Learning
  Heyan Chai (HarBin Institute of technology, ShenZhen); Zeyu Liu (HarBin Institute of technology, ShenZhen); Yongxin Tong (Beihang University); Ziyi Yao (Harbin Institute of technology, ShenZhen); Binxing Fang (Chinese Academy of Engineering); Qing Liao (Harbin Institute of Technology (Shenzhen))
- Hybrid Evaluation for Occlusion-based Explanations on CNN Inference Queries
  Guangyao Ding (East China Normal University); Chen Xu (East China Normal University); Weining Qian (East China Normal University)

Poster Session - [In Speys]

Posters of the 17:00 Session

Poster Session - [In B2B]

Posters of the 17:00 Session

Tencent Cloud

Tencent Cloud Database

Empowering Enterprises to Unlock Data Productivity

500k+ customers

50M+ cores

up to **80%** cost savings

Relational

TDSQL	TDSQL-C	TDSQL HTAP
Distributed Database	Cloud Native Database	Plugable HTAP Engine

MySQL · MariaDB · SQL Server · PostgreSQL

3rd-party Cloud Database

Non-Relational

VectorDB	KeeWiDB	TcaplusDB	CTSDB	KonisGraph
Vector Database	Key Value Database	Database for Gaming	Time-series Database	Graph Database

Redis · MongoDB · Memcached

3rd-party Cloud Database

DTS-DBbridge
Data Transfer & Migration Service

DBbrain
Database Autonomous Service

DMC
Database Management Service

TDSQL

#1 TPC-C | Distributed | Cost-Effective

01
814 million
transactions/min

02
99.999%
high availability

03
18s
or less in RTO

Contact us now to get started with Tencent Cloud
www.tencentcloud.com

Wednesday May 15th, 2024 16:30-17:00

Coffee Break - [In Speys]

Wednesday May 15th, 2024 17:00-18:30

Text, Semi-Structured Data, IR, Image, and Multimedia databases I - [In Theater 12, Chair: Zheng Wang]

- Alleviating the Inconsistency of Multimodal Data in Cross-Modal Retrieval
  Tieying Li (Northeastern University); Xiaochun Yang (Northeastern University); Yiping Ke (Nanyang Technological University); Bin Wang (Northeastern University); Yinan Liu (Northeastern University); Jiaxing Xu (NTU)
- Firzen: Firing Strict Cold-Start Items with Frozen Heterogeneous and Homogeneous Graphs for Recommendation
  Hulingxiao He (Peking University); Xiangteng He (Peking University); Yuxin Peng (Peking University); Zifei Shan (WeChat, Tencent); Xin Su (Tencent)
- Reconsidering Tree-based Methods for k-Maximum Inner-Product Search: The LRUS-CoverTree
  Hengzhao Ma (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences); Jianzhong Li (Harbin Institute of Technology); Yong Zhang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences)
- Cross-insight Trader: A Trading Approach Integrating Policies with Diverse Investment Horizons for Portfolio Management
  Zetao Zheng (University of Electronic Science and Technology of China); Jie Shao (University of Electronic Science and Technology of China); Shilong Deng (University of Electronic Science and Technology of China); Anjie Zhu (University of Electronic Science and Technology of China); Heng Tao Shen (University of Electronic Science and Technology of China (UESTC)); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- Unsupervised Multimodal Graph Contrastive Semantic Anchor Space Dynamic Knowledge Distillation Network For Cross-Media Hash Retrieval
  Yang Yu (Beijing University of Posts and Telecommunications); Meiyu Liang (Beijing University of Posts and Telecommunications); Mengran Yin (Beijing University of Posts and Telecommunications); Kangkang Lu (Beijing University of Posts and Telecommunications); Junping Du (Beijing University Of Posts And Telecommunications); Zhe Xue (Beijing University of Posts and Telecommunications)

Graphs, Networks, and Semistructured Data IV - [In Theater 11, Chair: Da Yan]

- AdaFGL: A New Paradigm for Federated Node Classification with Topology Heterogeneity
  Xunkai Li (Beijing Institute of Technology); Zhengyu Wu (Beijing Institute of Technology); Wentao Zhang (Mila); Henan Sun (Beijing Institute of Technology); Ronghua Li (Beijing Institute of Technology); Guoren Wang (Beijing Institute of Technology)
- Positive Communities on Signed Graphs That Are Not Echo Chambers: A Clique-Based Approach
  Alexander Zhou (Hong Kong University of Science and Technology); Yue Wang (Shenzhen Institute of Computing Sciences); Lei Chen (Hong Kong University of Science and Technology); Tamer Özsu (University of Waterloo)
- Maximal Biclique Enumeration: A Prefix Tree Based Approach
  Jiujian Chen (Beijing Institute of Technology); Kai Wang (Shanghai Jiao Tong University); Ronghua Li (Beijing Institute of Technology); Hongchao Qin (Beijing Institute of Technology); Xuemin Lin (University of New South Wales); Guoren Wang (Beijing Institute of Technology)
- Batch Hop-Constrained s-t Simple Path Query Processing in Large Graphs
  Kongzhang Hao (University of New South Wales); Long Yuan (Nanjing University of Science and Technology); Wenjie Zhang (University of New South Wales); Xuemin Lin (Shanghai Jiaotong University)
- On Searching Maximum Directed (k,l)-Plex
  Shuohao Gao (Harbin Institute of Technology, Shenzhen); Kaiqiang Yu (Nanyang Technological University); Shengxin Liu (Harbin Institute of Technology, Shenzhen); Cheng Long (Nanyang Technological University); Zelong Qiu (Harbin Institute of Technology, Shenzhen)

Explainability, Fairness, and Trust in Data Systems and Analysis III - [In Theater 10, Chair: George Fletcher]

- FAIRGEN: Towards Fair Graph Generation
  Lecheng Zheng (University of Illinois at Urbana-Champaign); Dawei Zhou (Virginia Tech); Hanghang Tong (University of Illinois at Urbana-Champaign); Jiejun Xu (HRL Laboratories, LLC); Yada Zhu (IBM); Jingrui He (University of Illinois at Urbana-Champaign)
- Fast, Robust and Interpretable Participant Contribution Estimation for Federated Learning
  Yong Wang (Tsinghua University); Yuyu Luo (HKUST (GZ)); Kaiyu Li (Tsinghua University); Guoliang Li (Tsinghua University); Yunyan Guo (Tsinghua University); Zhuo Wang (Tsinghua University)
- Exploring Optimal Parameters for Expected Results on Radius-Bounded k-Core Queries
  Chuanyu Zong (Shenyang Aerospace University); Zefang Dong (Shenyang Aerospace University); Xiaochun Yang (Northeastern University); Bin Wang (Northeastern University); Huaijie Zhu (Sun Yat-Sen University); Tao Qiu (Shenyang Aerospace University); Rui Zhu (Shenyang Aerospace University)
- Explaining Entity Matching with Clusters of Words
  Riccardo Benassi (Università di Modena e Reggio Emilia); Francesco Guerra (University of Modena e Reggio Emilia); Matteo Paganelli (Hasso Plattner Institute); Donato Tiano (Università degli Studi di Modena e Reggio Emilia)
- Fair Top-k Query on Alpha-Fairness
  Hao Liu (Hong Kong University of Science and Technology); Zheng Zhang (Individual); Raymond Chi-Wing Wong (Hong Kong University of Science and Technology); Min Xie (Shenzhen Institute of Computing Sciences); Bo Tang (Southern University of Science and Technology)

Data Mining and Knowledge Discovery X - [In Theater 9, Chair: Zeyi Wen]

- Enhancing the Performance of Bandit-Based Hyperparameter Optimization
  Yile Chen (SCUT); Jian Chen ("South China University of Technology, China"); Zeyi Wen (Hong Kong University of Science and Technology (Guangzhou)); Jin Huang (South China Normal Univesity)
- Unraveling the 'Anomaly' in Time Series Anomaly Detection: A Self-supervised Tri-domain Solution
  Yuting Sun (the University of Queensland); Guansong Pang (Singapore Management University); Guanhua Ye (Deep Neural Computing Company Limited); Tong Chen (The University of Queensland); Xia Hu (Rice University); Hongzhi Yin (The University of Queensland)
- A Robust Low-rank Tensor decomposition and Quantization based Compression Method
  Yudian Ouyang (Hunan University); kun xie (Hunan University); Jigang WEN (Hunan University of Science and Technology); Gaogang Xie (Computer Network Information Center, Chinese Academy of Sciences); Kenli Li (Hunan University)
- A Coarse-to-Fine Framework for Entity-Relation Joint Extraction
  Mingchen Zhang (Soochow University); Jiaan Wang (Soochow University); JIanfeng Qu (Soochow University); Zhixu Li (Fudan University); An Liu (Soochow University); Lei Zhao (Soochow University); Zhigang Chen (iFLYTEK CO., LTD); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- KGLink: A column type annotation method that combines knowledge graph and pre-trained language model
  Yubo Wang (HKUST); Hao Xin (Hong Kong University of Science and Technology); Lei Chen (Hong Kong University of Science and Technology)

Crowdsourcing III - [In Theater 4, Chair: Wookey Lee]

- Cross Online Ride-sharing for Multiple-Platform Cooperations in Spatial Crowdsourcing
  Yurong Cheng (Beijing institute of technology); Zhaohe Liao (Shanghai Jiao Tong University); Xiaosong Huang (peking university); Yi Yang (Beijing Institute of Technology); Xiangmin Zhou (RMIT University); Ye Yuan (Beijing Institute of Technology); Guoren Wang (Beijing Institute of Technology)
- Cooperative Air-Ground Instant Delivery by UAVs and Crowdsourced Taxis
  Junhui Gao (Northwestern Polytechnical University); Qianru Wang (Xidian University); Xin Zhang (National University of Defense Technology); Juan Shi (Air Force Engineering University); Xiang Zhao (National University of Defense Technology); Qingye Han (Chongqing University); yan pan (nudt)
- Urban Sensing for Multi-destination Workers via Deep Reinforcement Learning
  Shuliang Wang (Beijing Institute of Technology); Song Tang (Beijing Institute of Technology); Sijie Ruan (Beijing Institute of Technology); Cheng Long (Nanyang Technological University); Yuxuan Liang (The Hong Kong University of Science and Technology (Guangzhou)); Qi Li (Beijing institute of technology); Ziqiang Yuan (Beijing Institute of Technology); Jie Bao (Microsoft Research Asia, China); Yu Zheng (JD)
- Semi-Asynchronous Online Federated Crowdsourcing
  Xiangping Kang (Shandong University); Guoxian Yu (Shandong University); Jun Wang (Shandong University); Hui Li (Shandong University); Carlotta Domeniconi (George Mason University); Qingzhong Li (" Shandong University, China")

Panel on Data Management in the Cloud: Trends and Directions [In Theater 3, Chair: C. Mohan]

Moderator:

- C. Mohan, Hong Kong Baptist University, China

Panelists:

- Hanuma Kodavalla, Microsoft, USA
- Guoliang Li, Tsinghua University, China (TBC)
- Yannis Papakonstantinou, Google, USA
- Jingren Zhou, Alibaba, China

Description:

This panel will focus on the trends and directions with respect to data management in the cloud. The panelists who represent major cloud software vendors like Alibaba, Google, Huawei, and Microsoft, as well as universities will discuss their organizations' past and current areas of focus. Both product work and research directions will be covered. We will discuss what significant modifications to on-prem DBMSs were needed for cloud deployment, and what some of the open issues are. Leveraging AI in various components of cloud DBMSs will also be an important topic of discussion. Industry analysts' estimates for market size and different vendors' market share will also be highlighted.

About the Panelists:

Dr. C. Mohan is currently a Distinguished Professor of Science at Hong Kong Baptist University, a Distinguished Visiting Professor at Tsinghua University, and a member of the Board of Governors of Digital University Kerala. He retired in 2020 from being an IBM Fellow at the IBM Almaden Research Center in Silicon Valley. He was an IBM researcher for 38.5 years in the database, blockchain, AI and related areas, impacting numerous IBM and non-IBM products, the research and academic communities, and standards, especially with his invention of the well-known ARIES family of database locking and recovery algorithms, and the Presumed Abort distributed commit protocol. This IBM (1997-2020), ACM (2002-) and IEEE (2002-) Fellow has also served as the IBM India Chief Scientist (2006-2009). In addition to receiving the ACM SIGMOD Edgar F. Codd Innovations Award (1996), the VLDB 10 Year Best Paper Award (1999) and numerous IBM awards, Mohan was elected to the United States and Indian National Academies of Engineering (2009). This Distinguished Alumnus of IIT Madras (1977) received his PhD at the University of Texas at Austin (1981). He is the inventor of 50 patents. During 2021, Mohan was the Shaw Visiting Professor at the National University of Singapore. Since 2016, Mohan has been a Distinguished Visiting Professor of China's prestigious Tsinghua University. In 2023, he was named a Distinguished Professor of Science of Hong Kong Baptist University. Mohan has served on the advisory board of IEEE Spectrum, and on numerous conference and journal boards. During most of 2022, he was a consultant at Google with the title of Visiting Researcher. He has also been a Consultant to the Microsoft Data Team in 2020. He has given talks in 43 countries. More information can be found in the Wikipedia page at https://bit.ly/CMwIkP and his homepage at https://bit.ly/CMoDUK

Hanuma Kodavalla is a Technical Fellow in the Azure Databases group at Microsoft where he has been for more than twenty years. He previously worked at Data General, Digital Equipment Corporation, Oracle, Sybase and Asera. For more than three decades, Hanuma worked on many aspects of Relational Database Systems and has been instrumental in architecting multiple commercial database systems for high performance and high availability. Hanuma received BTech in Electronics and Communications in 1981 from National Institute of Technology, Warangal, India, MTech in Computer Science in 1983 from Indian Institute of Technology, Chennai, India, and MS in Computer Science in 1988 from University of Massachusetts, Amherst, USA. He has a few publications in database conferences and many patents related to novel implementation techniques for online transaction processing and data warehousing in the areas of concurrency control, recovery, high-availability, query processing and security.

Guoliang Li is a tenured full Professor of Department of Computer Science at Tsinghua University. His research interests include database systems, autonomous databases, machine learning for databases, human-in-the-loop data management, large-scale data cleaning and integration. He was the General chair of SIGMOD 2021, Demo chair of VLDB 2021, Industry chair of ICDE 2021. He got VLDB 2017 Early Research Contribution Award, TCDE 2014 Early Career Award, Best of SIGMOD Papers 2023, SIGMOD 2023 Research Highlight Award. VLDB 2023 Best Industry Paper Runner-up, VLDB 2020 Best Papers, CIKM 2017 Best Paper Award, KDD 2018 Best Papers, ICDE 2018 Best Papers, DASFAA 2023 Best Paper Award, DASFAA 2014 Best Paper Runnerup, APWeb 2014 Best Paper Award, EDBT 2013 Similarity Join and Search Champion. He regularly served as (Senior) PC Member of SIGMOD, VLDB, ICDE, KDD, WWW. He was serving as associate editor for IEEE TKDE, VLDB Journal, and IEEE Data Engineering Bulletin.

Yannis Papakonstantinou is a Distinguished Engineer, working on Query Processing and GenAI, at Google Cloud. He is also an Adjunct Professor of Computer Science and Engineering at the University of California, San Diego, following many years of having been a UCSD regular faculty member. Previously, he was an architect in query processing & ETL at Databricks. Earlier, he was a Senior Principal Scientist at Amazon Web Services from 2018-2021 and was a consultant for AWS since 2016. He was the CEO and Chief Scientist of Enosys Software, which built and commercialized an early Enterprise Information Integration platform for structured and semi-structured data. The Enosys Software was OEMed and sold under the BEA Liquid Data and BEA AquaLogic brand names, eventually acquired in 2003 by BEA Systems. His R&D work has been mostly on query processing with a focus on querying semi-structured data. He has published over 120 research articles that have received over 20,000 citations. Yannis holds a Diploma of Electrical Engineering from the National Technical University of Athens, MS, and PhD in Computer Science from Stanford University (1997).

Jingren Zhou currently holds the position of Chief Technology Officer at Alibaba Cloud, where he plays a pivotal role in driving technology innovation and product development. His responsibilities also include leading the development of AI foundational models and their applications in various business scenarios at Alibaba Cloud. Before this role, he has led work to develop advanced techniques for personalized search, product recommendation, and advertising at Alibaba's e-commerce platform and Alipay's online payment platform. Prior to his time at Alibaba, he was a veteran at Microsoft, where he led Big Data, and AI research and development. His research interests span across cloud computing, databases, and large-scale machine learning systems. He holds a BS in Computer Science from the University of Science and Technology of China, and a PhD in Computer Science from Columbia University. He is a Fellow of IEEE.

Industry - Query Performance [In Theater 2, Chair: Peter Boncz]

- Invited Talk: How I Learned to Stop Worrying About Benchmarks
  Hannes Muhleisen (DuckDB Labs)
- Addressing the Nested Data Processing Gap: JSONiq Queries on Snowflake through Snowpark
  Dan Graur (ETH Zurich); Remo Röthlisberger (ETH Zürich); Adrian Jenny (ETH Zurich); Ghislain Fourny (ETH Zurich); Filip Drozdowski (Snowflake); Choden Konigsmark (Snowflake); Ingo Müller (ETH Zurich); Gustavo Alonso (ETHZ)
- Bwe-tree: an Evolution of Bw-tree on Fast Storage
  Rui Wang (Alibaba Group); xinjun Yang (Alibaba Group); Feifei Li (Alibaba Group); David B Lomet (retired); Xin Liu (Alibaba Inc); panfeng zhou (Alibaba); Yongxiang Chen (Alibaba Group); David Zhang (Alibaba Group); Jingren Zhou (Alibaba Group); Jiesheng Wu (Alibaba)
- Resource Allocation with Service Affinity in Large-Scale Cloud Environments
  Zuzhi Chen (ByteDance Inc.); Fuxin Jiang (ByteDance Inc.); Binbin Chen (ByteDance Inc.); Yu Li (Bytedance.Inc); Yunkai Zhang (UC Berkeley); Tieying Zhang (Bytedance); Chao Huang (ByteDance Inc.); Rui Yang (ByteDance Inc.); Fan Jiang (ByteDance Inc.); Jianjun Chen (Bytedance); Wu Xiang (ByteDance Inc.); Guozhu Cheng (ByteDance Inc.); Rui Shi (ByteDance Inc.); Ning Ma (Xi'an Jiaotong University); Wei Zhang (South China University of Technology)
- Online Index Recommendation for Slow Queries
  Gan Peng (East China Normal University); Peng Cai (East China Normal University); Kaikai Ye (East China Normal University); Kai Li (Meituan); Jinlong Cai (Meituan); Yufeng Shen (Meituan); Han Su (Meituan); Weiyuan Xu (Meituan)

Tutorial on Privacy-Aware Analysis based on Data Series (Part B) [In Theater 1]

Abstract:
Data that is recorded about the operations of an organization constitutes a valuable source of information for monitoring and improvement. Specific use cases include the assessment of compliance to legal regulations, the analysis of performance bottlenecks, or the optimization of resource utilization. In recent years, a plethora of algorithms for operational analysis using data series, summarized as process mining, have been developed to support these use cases. Data series often contain sensitive information, though, about the individuals that act as service consumers or service providers. Personal information is only partially hidden by obfuscation and pseudonymization and potential privacy breaches need to be prevented for ethical, legal, and economic reasons.
This tutorial is devoted to methods for privacy-aware analysis using data series. It covers essential notions, reviews privacy-disclosure attacks, and outlines techniques to give formal privacy guarantees while largely maintaining the data's utility for operational analysis. The discussion is structured by the adopted perspective on the privacy of individuals, and the degree to which a data series contains contextual information.

Presenters:

Stephan Fahrenkrog-Petersen is a research group lead at the Weizenbaum Institute, Germany. He holds a PhD from Humboldt-Universität zu Berlin. His research was published in the proceedings of the premier conferences in the field and in international journals, such as ACM TMIS, DKE, and KAIS. His work received the Distinguished Paper Award at CAiSE 2021 and the Best Student Paper Award at ICPM 2021.

Han van der Aa is a junior professor in the Data and Web Science Group at the University of Mannheim, Germany. He obtained a PhD from the Vrije Universiteit Amsterdam in 2018. His research interests include process modelling, process mining, natural language processing, and complex event processing. His work has been published in journals including IEEE TKDE, Information Systems, and Decision Support Systems and at the BPM, CAISE, ICPM, ICDE, and SIGMOD conferences.

Matthias Weidlich is a professor and Chair of Databases and Information Systems at Humboldt-Universität zu Berlin, Germany. Matthias' research focuses on process-oriented and event-based information systems. His results appear regularly in premier conferences (SIGMOD, VLDB, ICDE, IJCAI, AAAI, BPM, CAiSE) and journals (TKDE, Information Systems, VLDB Journal) in the field. He serves as Co-Editor-in-Chief for the Information Systems journal and is a member of the steering committee of the ACM DEBS conference series.

Poster Session - [In Speys]

Posters of the 10:00 Session

Poster Session - [In B2B]

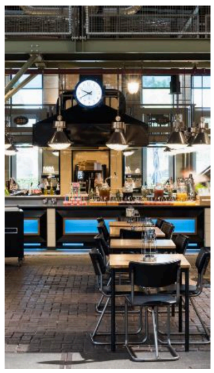
Posters of the 10:00 Session

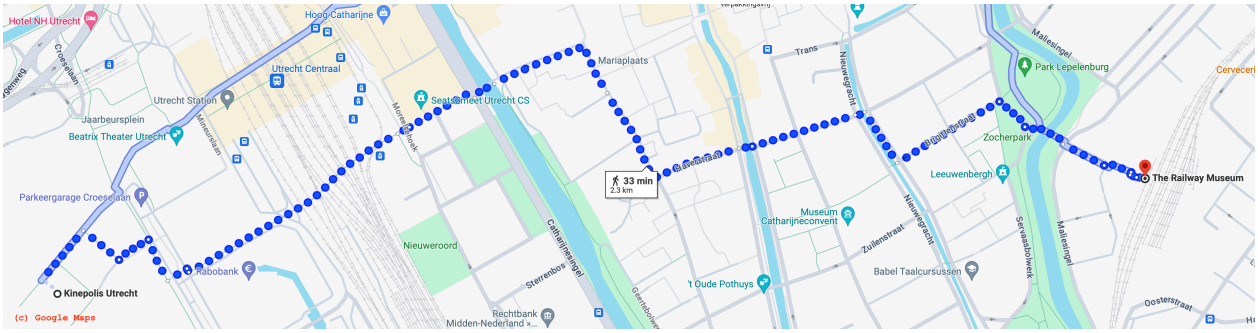
Wednesday May 15th, 2024 19:00-22:30

This year, the organization of the conference is offering two different types of banquet dinner. A dinner at the museum and a party-boat sail in the canals.

Dinner at the museum .... [At the Railway Museum (Spoorwegmuseum)]

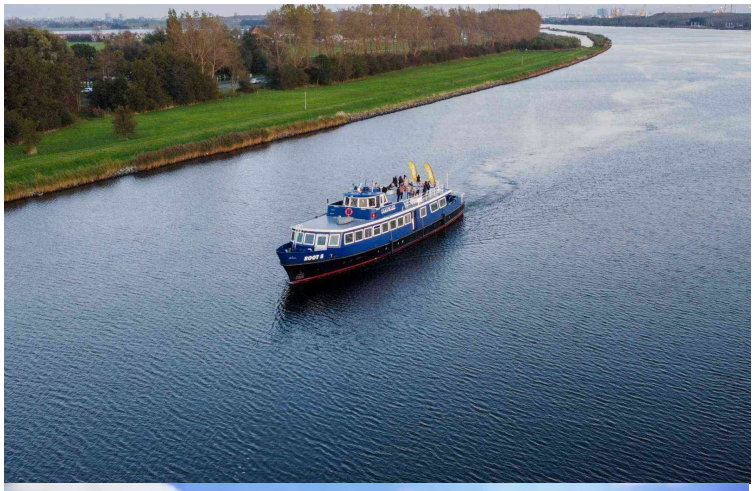
The ICDE24 Organization committee would like to ask the honor of your presence in a sit-down dinner at the Railway Museum (Spoorwegmuseum) in Utrecht. Come to have a dinner in a unique environment among the trains, while talking to fellow researchers, catching up with old connections, or making news. The museum is located at Maliebaanstation 16, 3581 XW Utrecht. Participants can arrive to the museum on foot from from Kinepolis. It takes only 30 min and the route is through the city, allowing you to admire its style and the lovely canal that goes though it.

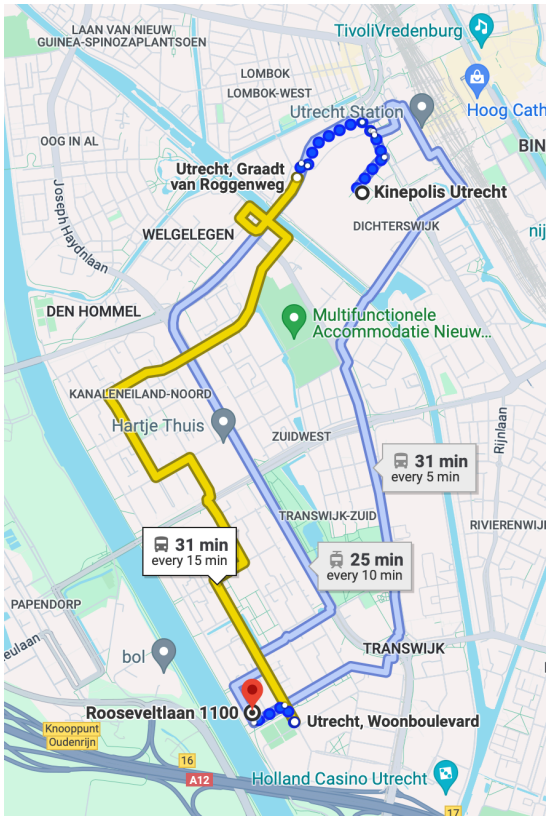




Dancing on the Canals ... [On Boat 8]

The organizing committee would like to invite you for a unique experience of a buffet (walking) dinner on the party-boat BOAT 8. Enjoy your dinner in the boat lounges or the deck outside under the music of the DJ ... and maybe dance as the boat sails through the canals. Transportation to and from the docks is done by public transportation. Participants are adviced to start leaving from the venue right after the end of the last presentation session of the day, and to have purchased their public transportation ticket in advance. This can be done at the the ticket machines at the train station. At the station there is also a ticket office (ov-chipkart). Instead of a two-way ticket it is also possible to buy a day-ticket with a minor difference in price. The dock where the boat leaves is at Rooseveltlaan 1100, 3526 BT Utrecht. To get there one can get from the train station the bus number 7 or the tram 20/21 that both leave from very close to Kinepolis (see map). In the case of the tram, you need to step out at station Vasco da Gamalaan and walk from there. In the case of the bus number 7, one needs to step out at station Woonboulevard.





Thursday May 16th, 2024 8:30-9:30

Keynote 3 [In Theater 12, Chair: Yannis Velegrakis]

AI Systems beyond Accelerating Linear Algebra
by Christos Kozyrakis (Stanford University).

Over the past decade, there has been remarkable progress in the co-design of hardware and software systems for artificial intelligence (AI). Much of this progress has focused on accelerating computationally-intensive operations, such as the matrix multiplications in AI training and inference tasks. This talk will address the broader systems issues that are now emerging as significant bottlenecks for AI workloads. We will review challenges such as making inference resource efficient, optimizing workloads involving multiple AI tasks, and feeding AI workloads with data. We will advocate for the design of AI infrastructure that looks more like the scale-out systems used for cloud computing, rather than the supercomputing systems used for HPC. Finally, we will underscore the need to broaden the scope of AI systems co-design to encompass the applications themselves.

Christos Kozyrakis is a Professor of Electrical Engineering and Computer Science at Stanford University. He is also the faculty director of the Stanford Platform Lab. Christos' research focuses on computer architecture and systems software. He is currently working on cloud computing technology, systems design for artificial intelligence, and artificial intelligence for systems design. Christos holds a BS degree from the University of Crete (Greece) and a PhD degree from the University of California at Berkeley (USA). He is a fellow of the ACM and the IEEE. He has received the ACM SIGARCH Maurice Wilkes Award, the ISCA Influential Paper Award, the ASPLOS Influential Paper Award the NSF Career Award, the Okawa Foundation Research Grant, and faculty awards by IBM, Microsoft, and Google. Christos has also worked for technology companies such as Google and Intel and has helped launch AI infrastructure startups such as Enfabrica and Plix.

Thursday May 16th, 2024 9:30-10:00

Coffee Break - [In Speys]

Thursday May 16th, 2024 10:00-11:50

Future Technologies 1 - [In Theater 12, Chair: Beng Chin Ooi]

- SQL++: We Can Finally Relax!.
  by Michael Carey, Don Chamberlin, Almann Goo, Kian Win Ong, Yannis Papakonstantinou, Chris Suver, Sitaram Vemulapalli, and Till Westmann
- Data Flow Architectures for Data Processing on Modern Hardware.
  by Alberto Lerner (University of Fribourg), Gustavo Alonso (ETHZ)
- Personal Manifold: Management of Personal Data in the Age of Large Language Models.
  by Yuliang Li (Meta), Alon Halevy (Meta), Wang-Chiew Tan (Meta)
- Applications and Challenges for Large Language Models: From Data Management Perspective.
  by Meihui Zhang (Beijing Institute of Technology), Zhaoxuan Ji (Beijing Institute of Technology), ZHAOJING LUO (Beijing Institute of Technology), Yuncheng Wu (Renmin University of China), Chengliang Chai (Beijing Institute of Technology)
- Routing with Massive Trajectory Data
  by C. S. Jensen (Aalborg University), B. Yang, C. Guo, J. Hu.
- When Data Pricing Meets Non-cooperative Game Theory.
  by Jinfei Liu (Zhejiang University), Yuran Bi (Zhejiang University), Yihang Wu (Zhejiang University), Kui Ren (Zhejiang University), Li Xiong (Emory University).

Thursday May 16th, 2024 11:50-13:15

Lunch - [In Speys]

Thursday May 16th, 2024 13:15-14:45

Database technology for Blockchains II - [In Theater 12, Chair: Jiang Xiao]

- Porygon: Scaling Blockchain via 3D Parallelism
  Wuhui Chen (Sun Yat-sen University); Ding Xia (Sun Yat-sen University); Zhongteng Cai (The Ohio State University); Hong-Ning Dai (Hong Kong Baptist University); Jianting Zhang (Purdue University); Zicong Hong (The Hong Kong Polytechnic University); Junyuan Liang (Sun Yat-sen University); Zibin Zheng (Sun Yat-sen University)
- Authenticated Keyword Search on Large-scale Graphs in Hybrid-Storage Blockchains
  Siyu Li (Beijing Institute of Technology); Zhiwei Zhang (Beijing Institute of Technology); Jiang Xiao (Huazhong University of Science and Technology); Meihui Zhang (Beijing Institute of Technology); Ye Yuan (Beijing Institute of Technology); Guoren Wang (Beijing Institute of Technology)
- MuFuzz: Sequence-Aware Mutation and Seed Mask Guidance for Blockchain Smart Contract Fuzzing
  Peng Qian (Zhejiang University); Hanjie Wu (Zhejiang Gongshang University); Du ZeRen (Zhejiang Gongshang University); Turan Vural (UCLA); Dazhong Rong (Zhejiang University); Zheng Cao (Zhejiang University); Zhang Lun (Goplus Security); Yanbin Wang (Zhejiang University); Jianhai Chen (Zhejiang University); Qinming He (Zhejiang University)
- Authenticated Subgraph Matching in Hybrid-Storage Blockchains
  Siyu Li (Beijing Institute of Technology); Zhiwei Zhang (Beijing Institute of Technology); Meihui Zhang (Beijing Institute of Technology); Ye Yuan (Beijing Institute of Technology); Guoren Wang (Beijing Institute of Technology)
- Authenticating Multi-Chain Queries: Verifiable Virtual Filesystem Is All You Need
  Haixin Wang (Hong Kong Baptist University); Cheng Xu (Hong Kong Baptist University); Xiaojie Chen (Ant Group); Ce Zhang (Hong Kong Baptist University); Haibo Hu (Hong Kong Polytechnic University); Shikun Tian (Ant Group); Ying Yan (Ant Group); Jianliang Xu (Hong Kong Baptist University)

Modern Hardware and In-Memory Database Systems I - [In Theater 11, Chair: Tilmann Rabl]

- Enabling Efficient NVM-Based Text Analytics without Decompression
  Xiaokun Fang (Renmin University of China); Feng Zhang (Renmin University of China); Junxiang Nong (Renmin University of China); Mingxing Zhang (Tsinghua University); Puyun Hu (Renmin University of China); yunpeng chai (renmin university of china); Xiaoyong Du (Renmin University of China)
- F-TADOC: FPGA-Based Text Analytics Directly on Compression with HLS
  Yanliang Zhou (Renmin University of China); Feng Zhang (Renmin University of China); Tuo Lin (Advanced Micro Device); Yuanjie Huang (Advanced Micro Device); Saiqin Long (Jinan University); Jidong Zhai (Tsinghua University); Xiaoyong Du (Renmin University of China)
- Robust External Hash Aggregation in the Solid State Age
  Laurens Kuiper (CWI); Peter Boncz (CWI); Hannes Mühleisen (Centrum Wiskunde & Informatica)
- Neos: A NVMe-GPUs Direct Vector Service Buffer in User Space
  Yuchen Huang (East China Normal University); Xiaopeng Fan (East China Normal University); song yan (华东师范大学); Chuliang Weng (East China Normal University)
- TEngine: A Native Distributed Table Storage Engine
  Xiaopeng Fan (East China Normal University); song yan (华东师范大学); Yuchen Huang (East China Normal University); Chuliang Weng (East China Normal University)

Graphs, Networks, and Semistructured Data V - [In Theater 10, Chair: Xiangyu KE]

- Masked Graph Modeling with Multi-View Contrast
  Yanchen Luo (University of Science and Technology of China); Sihang Li (University of Science and Technology of China); Yongduo Sui (University of Science and Technology of China); Junkang Wu (University of Science and Technology of China); Jiancan Wu (University of Science and Technology of China); Xiang Wang (National University of Singapore)
- Multi-View Teacher with Curriculum Data Fusion for Robust Unsupervised Domain Adaptation
  Yuhao Tang (Peking University); Junyu Luo (Peking University); Ling Yang (Peking University); Xiao Luo (UCLA); Wentao Zhang (Peking University); Bin Cui (Peking University)
- GSHOP: Towards Flexible Pricing for Graph Statistics
  Chen Chen (Beijing Institute of Technology); Ye Yuan (Beijing Institute of Technology); ZHENYU WEN (Zhejiang University of Technology); Yu-Ping Wang (Beijing Institute of Technology); Guoren Wang (Beijing Institute of Technology)
- LearnSC: An Efficient and Unified Learning-based Framework for Subgraph Counting Problem
  Wenzhe Hou (National University of Defense Technology); Xiang Zhao (National University of Defense Technology); Bo Tang (Southern University of Science and Technology)
- AFTER: Adaptive Friend Discovery for Temporal-spatial and Social-aware XR
  Bing-Jyue Chen (Academia Sinica); Chiok Yew Ho (National Taiwan University); De-Nian Yang (Academia Sinica)

Graphs, Networks, and Semistructured Data VI - [In Theater 9, Chair: Tangwen Qian]

- Reducing Resource Usage for Continuous Model Updating and Predictive Query Answering in Graph Streams
  Qu Liu (University of Massachusetts, Lowell); Adam King (University of Massachusetts Lowell); Tingjian Ge (University of Massachusetts, Lowell)
- Graph Anomaly Detection with Domain-agnostic Pre-training and Few-shot Adaptation
  Xujia Li (The Hong Kong University of Science and Technology); Lei Chen (Hong Kong University of Science and Technology)
- NC-ALG: Graph-based Active Learning under Noisy Crowd
  Wentao Zhang (Peking University); YeXin Wang (Peking University); Zhenbang You (Stanford University, Peking University); Yang Li (Tencent Inc.); Gang Cao (Beijing Academy of Artificial Intelligence); Zhi Yang (Peking University); Bin Cui (Peking University)
- Fast Multilayer Core Decomposition and Indexing
  Dandan Liu (Harbin Institute of Technology); Run-An Wang (Harbin Institute of Technology); Zhaonian Zou (Harbin Institute of Technology); Xin Huang (Hong Kong Baptist University)
- CINA: Curvature-based Integrated Network Alignment with Hypergraph
  Pengfei Jiao (Hangzhou Dianzi University); Yuanqi Liu (Hangzhou Dianzi University); Yinghui Wang (Tianjin University); Ge Zhang (Macquarie University)

PhD Symposium - [In Theater 4, Chair: Panos K. Chrysanthis]

- PR-GNN: Enhancing PoC Report Recommendation with Graph Neural Network
  Jiangtao Lu (Army Engineering University of PLA)
- Cascade: Optimal Transaction Scheduling for High-Contention Workloads
  Tim Baccaert (Vrije Universiteit Brussel)
- Construction and enhancement of an RNA-based knowledge graph for discovering new RNA drugs
  Emanuele Cavalleri (AnacletoLab, University of Milan)
- Enhancing Data Systems Performance by Exploiting SSD Concurrency & Asymmetry
  Tarikul Islam Papon (Boston University)
- Differential Analysis for System Provenance
  Yuta Nakamura (DePaul University)

Lightning Talks - [In Theater 3, Chair: Nikos Ntarmos]

- Evaluating Text-to-SQL Models on Real-world Enterprise Data
  Manasi Ganti (University of Washington Seattle); Laurel Orr (Numbers Station AI); Sen Wu (Numbers Station AI)
- Synergies between Graph Data Management and Machine Learning in Graph Data Science Pipeline
  Arijit Khan (Aalborg University)
- Large Language Models as Storage for SQL Querying
  Paolo Papotti (EURECOM)
- Accelerating Deletion Interventions on OLAP workload
  Haneen Mohammed (Columbia University); Alexander Yao (Columbia University); Lampros Flokas (Columbia University); Zhong Hongbin (Northeastern University); Charlie G Summers (Columbia University); Eugene Wu (Columbia University)
- User Learning In Interactive Data Exploration
  Nischal Aryal (Oregon State University); Sanad Saha (Oregon State University); Leilani Battle (University of Washington); Arash Termehchy (Oregon State University)
- Multivariate similarity search -- a call for a new breed of similarity search algorithms
  Odysseas Papapetrou (TU Eindhoven); Jens d'Hondt (Eindhoven University of Technology)
- Towards Streaming Consistency Management
  Samuele Langhi (Lyon 1 University); Angela Bonifati (Univ. of Lyon); Riccardo Tommasini (INSA Lyon - LIRIS)
- Unveiling Dis-Integration
  George Papadakis (University of Athens); Ekaterini Ioannou (Tilburg University); Yannis Velegrakis (Utrecht University and University of Trento)
- Cross-Source ML Model Training
  wenbo sun (Technical University of Delft); Rihan Hai (TU Delft)
- Why Model-Based Lossy Compression is Great for Wind Turbine Analytics
  Søren Kejser Jensen (Aalborg University); Christian Thomsen (Aalborg University); Torben Bach Pedersen (Aalborg University)*; Carlos Enrique Muniz Cuza (Aalborg University); Abduvoris Abduvakhobov (Aalborg University)
- Towards Explainability in Retrieval-Augmented LLMs
  Joel E Rorseth (University of Waterloo); Parke Godfrey (York University); Lukasz Golab (University of Waterloo); Divesh Srivastava (AT&T Chief Data Office); Jaroslaw Szlichta (York University and IBM CAS)
- Benchmarking Data Management Systems for Microservices
  Rodrigo Nunes Laigner (University of Copenhagen); Yongluan Zhou (University of Copenhagen)
- Exploring the Space of Model Comparisons
  Andy Yu (Waterloo University); Parke Godfrey (York University); Lukasz Golab (University of Waterloo); Divesh Srivastava (AT&T Chief Data Office); Jaroslaw Szlichta (York University and IBM CAS)
- On Native Location-Optimized Data Systems
  Walid G Aref (Purdue)
- Observations and Opportunities in Solving Large-Scale Graph Data Processing Challenges at ByteDance by Using Heterogeneous Hardware
  cheng chen (bytedance); Shuai Zhang (Bytedance)

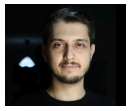
Industry - Tuning and Industrial Applications - [In Theater 2, Chair: Yingqiang Zhang]

- On Tuning Raft for IoT Workload in Apache IoTDB
  Tian Jiang (Tsinghua University); Xiangdong Huang (Tsinghua University); Shaoxu Song (Tsinghua University); Chen Wang (Timecho Limited); Jianmin Wang ("Tsinghua University, China")
- Enabling Roll-up and Drill-down Operations in News Exploration with Knowledge Graphs for Due Diligence and Risk Management
  Sha Wang (Singapore Management University); Yuchen Li (Singapore Management University); Hanhua Xiao (Singapore Management University); Zhifeng Bao (RMIT University); Lambert Deng (DBS); Yanfei Dong (PayPal)
- Multifaceted Reformulations for Null \& Low queries and its parallelism with Counterfactuals
  Jayanth Yetukuri (UC Santa Cruz); yuyan wang (ebay); Ishita K Khan (ebay); liyang hao (ebay); zhe wu (ebay); Yang Liu (UC Santa Cruz)
- An Effective, Efficient, and Stable Framework for Query Clustering
  Liuqing Li (Yahoo!); Chang Lu (Stevens Institute of Technology); Donghyun Kim (Yahoo!); Xinyue Wang (Yahoo!); Rao Shen (Yahoo!)
- A Framework for Continuous kNN Ranking of EV Chargers with Estimated Components
  Soteris Constantinou (University of Cyprus); Constantinos Costa (Rinnoco Ltd); Andreas Konstantinidis (Frederick University); Mohamed Mokbel (University of Minnesota - Twin Cities); Demetrios Zeinalipour-Yazti (University of Cyprus)

Tutorial on Robust Query Optimization in the Era of Machine Learning: State-of-the-Art and Future Directions [In Theater 1]

Abstract:
Query optimizers are an essential component of relational database management systems (RDBMSs) as they search for an execution plan that is expected to be optimal for a given query. However, they commonly use parameter estimates that are often inaccurate and make assumptions that may not hold in practice. Consequently, the optimizer may select suboptimal execution plans at runtime, when these estimates and assumptions are not valid, which may result in poor query performance. Therefore, query optimizers do not sufficiently support the robustness of the database system. In this tutorial we aim to explore the notion of robustness of a query execution plan, as well as how robustness is evaluated or even further supported. Firstly, we provide a comprehensive definition of robustness in this context. Next, we review the approaches proposed in the literature to address the issue of robustness, including techniques that rely on query re-optimization, discovering parameters, quantifying robustness, as well as recent techniques that employ machine learning. We emphasize the comparison of traditional cost-model-based and recent ML-based techniques concerning their capacity to address the issue of robustness in query optimization. Finally, we discuss the limitations and gaps in the current literature and provide some recommendations for future research directions.

Presenters:



Amin Kamali is a Ph.D. Candidate in Digital Transformation and Innovation at the University of Ottawa. Over the past decade, he has occupied diverse roles at IBM, all centered around Data and AI. These roles have spanned a spectrum from Business Intelligence to Digital Transformation, Data Science and Machine Learning. Amin holds a B.Sc. in Industrial Engineering from Sharif University of Technology and an M.Sc. in Systems Science from the University of Ottawa. His primary research interest revolves around delving into the latest AI breakthroughs and their potential to transform data systems. Amin has been privileged to organize multiple workshops for various audiences and to present at several academic conferences.

Verena Kantere has held academic positions as: Professor at the School of EECS of the University of Ottawa, Assistant Professor at the School of ECE of the National Technical University of Athens (NTUA), Maître d'Enseignement et de Recherche at the Centre Universitaire d' Informatique of the University of Geneva and Junior Assistant Professor at the Department of Electrical Engineering and Information Technology at the Cyprus University of Technology. She has conducted research for many years in the domain of data management, showing results in Peer-to-Peer systems, scientific data management, cloud data management, Big Data management, and analysis. She has received an Engineering Diploma and a Ph.D. from the NTUA and a M.Sc. from the University of Toronto.

Calisto Zuzarte is a Senior Technical Staff Member (STSM) in the Db2 development organization in IBM. His expertise is in database query optimization and has 60+ patents and 60+ research publications related to this area. His current interest is in the application of machine learning in query optimization and optimization in the Lakehouse environment.

Poster Session - [In Speys]

Posters of the 16:30 Session

Poster Session - [In B2B]

Posters of the 16:30 Session

Thursday May 16th, 2024 15:00-16:00

Graphs, Networks, and Semistructured Data VII - [In Theater 12, Chair: Xin Huang]

- Open-World Semi-Supervised Learning for Node Classification
  Yanling Wang (Zhongguancun Laboratory); Jing Zhang (Renmin University of China); LingXi Zhang (Renmin University of China); Lixin Liu (Alibaba Group); Yuxiao Dong (Tsinghua); Cuiping Li (Renmin University of China); Hong Chen (" Renmin University, China"); Hongzhi Yin (The University of Queensland)
- Scalable Community Search with Accuracy Guarantee on Attributed Graphs
  Yuxiang Wang (Hangzhou Dianzi University); ShuZhan Ye (Hangzhou Dianzi University); Xiaoliang Xu (Hangzhou Dianzi University); Yuxia Geng (Hangzhou Dianzi University); Zhenghe Zhao (Hangzhoudianzi University); Xiangyu Ke (Zhejiang University, China); Tianxing Wu (Southeast University)
- From Motif to Path: Connectivity and Homophily
  Qihao Wang (University of Illinois at Urbana Champaign); Hongtai Cao (University of Illinois at Urbana-Champaign); Xiaodong Li (The University of Hong Kong); Kevin Chen-Chuan Chang (University of Illinois at Urbana-Champaign); Reynold Cheng ("The University of Hong Kong, China")

Modern Hardware and In-Memory Database Systems II - [In Theater 11, Chair: Amir Shaikhha]

- DmRPC: Disaggregated Memory-aware Datacenter RPC for Data-intensive Applications
  Jie Zhang (Zhejiang University); Xuzheng Chen (Zhejiang University); Yin Zhang (Zhejiang University); Zeke Wang (Zhejiang University)
- RapidGKC: GPU-accelerated K-mer Counting
  Yiran Cheng (The Hong Kong University of Science and Technology); Xibo Sun (Hong Kong University of Science and Technology); Qiong Luo (Hong Kong University of Science and Technology)
- Sylvie: 3D-adaptive and Universal System for Large-scale Graph Neural Network Training
  Meng Zhang (Nanyang Technological University); Qinghao Hu (Nanyang Technological University); Cheng Wan (Georgia Institute of Technology); Haozhao Wang (Nanyang Technological University); Peng Sun (SenseTime Group Ltd); Yonggang Wen (Nanyang Technological University); Tianwei Zhang (Nanyang Technological University)

Graphs, Networks, and Semistructured Data VIII - [In Theater 10, Chair: Zhenhua Huang]

- Self-Training GNN-based Community Search in Large Attributed Heterogeneous Information Networks
  Yuan Li (North China University of Technology); Xiuxu Chen (North China University of Technology); Yuhai Zhao (Northeastern University); Wen Shan (Singapore University of Social Sciences); Zhengkui Wang (Singapore Institute of Technology); Guoli Yang (Advanced Institute of Big Data, Beijing); Guoren Wang (Beijing Institute of Technology)
- HGAMLP: Heterogeneous Graph Attention MLP with De-redundancy Mechanism
  Yuxuan Liang (Peking University); Wentao Zhang (Peking University); Zeang Sheng (Peking University); Ling Yang (Peking University); Jiawei Jiang (Wuhan University); Yunhai Tong (Peking University); Bin Cui (Peking University)
- FocusCore Decomposition of Multilayer Graphs
  Run-An Wang (Harbin Institute of Technology); Dandan Liu (Harbin Institute of Technology); Zhaonian Zou (Harbin Institute of Technology)

Graphs, Networks, and Semistructured Data IX - [In Theater 9, Chair: Wentao Li]

- Search to Fine-tune Pre-trained Graph Neural Networks for Graph-level Tasks
  Zhili WANG (HKUST); Shimin Di (The Hong Kong University of Science and Technology); Lei Chen (Hong Kong University of Science and Technology); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- BOURNE: Bootstrapped Self-supervised Learning Framework for Unified Graph Anomaly Detection
  Jie Liu (Northwestern Polytechnical University); Mengting He (Northwestern Polytechnical University); Xuequn Shang (Northwestern Polytechnical University); Jieming Shi (The Hong Kong Polytechnic University); Bin Cui (Peking University); Hongzhi Yin (The University of Queensland)
- Discovering Personalized Characteristic Communities in Attributed Graphs
  Yudong Niu (Singapore Management University); Yuchen Li (Singapore Management University); Panagiotis Karras (University of Copenhagen); Yanhao Wang (East China Normal University); Zhao Li (Hangzhou Yugu Technology)

Data Stream Systems and Edge Computing I - [In Theater 4, Chair: Ming Hu]

- Fast Parallel Recovery for Transactional Stream Processing on Multicores
  Jianjun Zhao (Huazhong University of Science and Technology); Haikun Liu (Huazhong University of Science and Technology); Shuhao Zhang (Nanyang Technological University); Zhuohui Duan (Huazhong University of Science and Technology); Xiaofei Liao (HUST); Hai Jin (Huazhong University of Science and Technology); Yu Zhang (Huazhong University of Science and Technology)
- PP-Stream: Toward High-Performance Privacy-Preserving Neural Network Inference via Distributed Stream Processing
  Qingxiu Liu (Peking University and The Chinese University of Hong Kong); Qun Huang (Peking University); Xiang Chen (Zhejiang University); Sa Wang (Chinese Academy of Sciences); Wenhao Wang (Institute of Information Engineering, CAS); Shujie Han (Peking University); Patrick P. C. Lee (The Chinese University of Hong Kong)
- AdaEdge: A Dynamic Compression Selection Framework for Resource Constrained Devices
  Chunwei Liu (MIT); John Paparrizos (The Ohio State University); Aaron J Elmore (University of Chicago)

Data Mining and Knowledge Discovery XI - [In Theater 3, Chair: Ning Liu]

- Learning k-Determinantal Point Processes for Personalized Ranking
  Yuli LIU (ANU); Lexing Xie (Australian National University); Christian Walder (Google Deepmind)
- A Unified Replay-based Continuous Learning Framework for Spatio-Temporal Prediction on Streaming Data
  Hao Miao (Aalborg University); Yan Zhao (Aalborg University); Chenjuan Guo (ECNU); Bin Yang (Aalborg University); Kai Zheng (University of Electronic Science and Technology of China); Feiteng Huang (Huawei Cloud Database Innovation Lab); jiandong xie (HUAWEI TECHNOLOGIES CO.LTD.); Christian S. Jensen (Aalborg University)
- Representation Learning of Tangled Key-Value Sequence Data for Early Classification
  Tao Duan (Xi'an Jiaotong University); Junzhou Zhao (Xi'an Jiaotong University); Shuo Zhang (Xi'an Jiaotong University); Jing Tao (Xi'an Jiaotong University); Pinghui Wang (Xi'an Jiaotong University)

Database Security and Privacy I - [In Theater 2, Chair: TBD]

- LDPRecover: Recovering Frequencies from Poisoning Attacks against Local Differential Privacy
  Xinyue Sun (Beihang university); Qingqing Ye (Hong Kong Polytechnic University); Haibo Hu (Hong Kong Polytechnic University); Jiawei Duan (The Hong Kong Polytechnic University); Tianyu Wo (Beihang University); Jie Xu (University of Leeds); Renyu Yang (Beihang University)
- Differentially Private Graph Neural Networks for Link Prediction
  Xun Ran (The Hong Kong Polytechnic University (PolyU)); Qingqing Ye (Hong Kong Polytechnic University); Haibo Hu (Hong Kong Polytechnic University); Xin Huang (Hong Kong Baptist University); Jianliang Xu (Hong Kong Baptist University); Jie Fu (East China Normal University)
- Secure and Practical Functional Dependency Discovery in Outsourced Databases
  Xinle Cao (Zhejiang University); Yuhan Li (Zhejiang University); Dmytro Bogatov (Amazon); Jian Liu (Zhejiang University); Kui Ren (Zhejiang University)

Data Mining and Knowledge Discovery XIII [In Theater 1, Chair: Yanchao Tan]

- Multi-Modality is All You Need for Transferable Recommender Systems
  Youhua Li (ShanghaiTech University); Hanwen Du (The Hong Kong University of Science and Technology); Yongxin Ni (Westlake University); Pengpeng Zhao (Soochow University); Fajie Yuan (Westlake University); Qi Guo (Institute of Computing technology); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- Multi-view Attentive Variational Learning for Group Recommendation
  wen yang (Soochow University); Jiajie Xu (Soochow University); Rui Zhou (Swinburne University of Technology); Lu Chen (Swinburne University of Technology); Jianxin Li (Deakin University); Pengpeng Zhao (Soochow University); Chengfei Liu (Swinburne University of Technology)
- Corruption Robust Dynamic Pricing in Liner Shipping under Capacity Constraint
  Yongyi Hu (Shanghai Jiao Tong University); Xueyan Li (Shanghai Jiao Tong University); Xikai Wei (Shanghai Jiao Tong University); Yangguang Shi (Shandong University); Xiaofeng Gao (Shanghai Jiaotong University); Guihai Chen (Shanghai Jiao Tong University)

Poster Session - [In Speys]

Posters of the 10:00 Session

Demonstrations - Group A - [In B2B]

- Entity/Relationship Profiling
  by Henning Koehler (Massey University), Sebastian Link (University of Auckland)
- GA-Tags: Data Enrichment with an Automatic Tagging System Utilizing Large Language Models
  by Genki Kusano (NEC)
- Comparing Personalized Relevance Algorithms for Directed Graphs
  by Luca Cavalcanti (University of Trento), Cristian Consonni (Joint Research Centre, European Commission), Martin MB Brugnara (University of Trento), David Laniado (Eurecat), Alberto Montresor (University of Trento)
- FSM-Explorer: An Interactive Tool for Frequent Subgraph Pattern Mining from a Big Graph
  by Jalal Khalil (St. Cloud State University), Da Yan (Indiana University Bloomington), Lyuheng Yuan (Indiana University Bloomington), Jiao Han (George Washington University), Saugat Adhikari (Indiana University Bloomington), Cheng Long (Nanyang Technological University), Yang Zhou (Auburn University)
- TASKS: A Real-Time Query System for Instant Error-Tolerant Spatial Keyword Queries on Road Networks
  by Chengyang Luo (Zhejiang University), lu jin (Zhejiang University), Qing Liu (Zhejiang University), Yunjun Gao (Zhejiang University), Lu Chen (Zhejiang University)
- VASIM: Vertical Autoscaling Simulator Toolkit
  by Anna Pavlenko (Microsoft Gray Systems Lab), Karla Saur (Microsoft), Yiwen Zhu (Microsoft), Brian Kroth (Microsoft), Joyce Cahoon (Microsoft), Jesús Camacho-Rodríguez (Microsoft)
- Demonstration of FeVisQA: Free-Form Question Answering over Data Visualization
  by Yuanfeng Song (The Hong Kong University of Science and Technology), 锦伟 卢 (深圳大学), Xuefang Zhao (WeBank Co., Ltd), Raymond Chi-Wing Wong (Hong Kong University of Science and Technology), Haodi Zhang (Shenzhen University)
- CleanEr: Interactive, Query-Guided Error Mitigation for Data Cleaning Systems
  by Ran Schreiber (Bar-Ilan University), Yael Amsterdamer (Bar-Ilan university)
- Wearables for Health (W4H) Toolkit for Acquisition, Storage, Analysis and Visualization of Data from Various Wearable Devices
  by Arash Hajisafi (University of Southern California), Maria Despoina Siampou (University of Southern California), Jize Bi (University of Southern California), Luciano Nocera (University of Southern California), Cyrus Shahabi (Computer Science Department. University of Southern California)

Demonstrations - Group C - [In B2B]

- Explaining Expert Search Systems with ExES
  by Kiarash Golzadeh (University of Waterloo), Lukasz Golab (University of Waterloo), Jaroslaw Szlichta (York University and IBM CAS)
- RAGE Against the Machine: Retrieval-Augmented LLM Explanations
  by Joel E Rorseth (University of Waterloo), Parke Godfrey (York University), Lukasz Golab (University of Waterloo), Divesh Srivastava (AT&T Chief Data Office), Jaroslaw Szlichta (York University and IBM CAS)
- FairCR - an evaluation and recommendation system for fair classification algorithms
  by Nico Lässig (University of Stuttgart), Melanie Herschel (Universität Stuttgart)
- GraphLingo: Domain Knowledge Exploration by Synchronizing Knowledge Graphs and Large Language Models
  by Duy Le (Case Western Reserve University), Kris Zhao (Case Western Reserve University), Mengying Wang (Case Western Reserve University), Yinghui Wu (Case Western Reserve University)
- MixedSearch: An Interactive System of Searching for the Best Tuple with Mixed Attributes
  by Weicheng Wang (Hong Kong University of Science and Technology), Min Xie (Shenzhen Institute of Computing Sciences), Raymond Chi-Wing Wong (Hong Kong University of Science and Technology)
- MorphStream: Scalable Processing of Transactions over Streams
  by Siqi Xiang (Singapore University of Technology and Design), Zhonghao Yang (Singapore University of Technology and Design), Shuhao Zhang (Nanyang Technological University), Jianjun Zhao (Huazhong University of Science and Technology), Yancan Mao (National University of Singapore)
- FONT: A Flexible Polystore Evaluation Platform
  by Gengyuan Shi (Tsinghua University), Chaokun Wang (Tsinghua University), Minghao Zhang (Tsinghua University), Binbin Wang (MSFT)
- CAMO: Explaining Consensus Across MOdels
  by Andy Yu (Waterloo University), Parke Godfrey (York University), Lukasz Golab (University of Waterloo), Divesh Srivastava (AT&T Chief Data Office), Jaroslaw Szlichta (York University and IBM CAS)
- Pyneapple-R: Scalable and Expressive Spatial Regionalization
  by Yunfan Kang (University of California Riverside), Yongyi Liu (University of California, Riverside), Hussah Alrashid (University of California Riverside), Akash Bilgi (University of California, Riverside), Siddhant Purohit (University of California, Riverside), Ahmed Mahmood (Google), Sergio Rey (San Diego State University), Amr Magdy (University of California Riverside)

Thursday May 16th, 2024 16:00-16:30

Coffee Break - [In Speys]

Graphs, Networks, and Semistructured Data X - [In Theater 12, Chair: Hanchen Wang]

- TP-GCN: Continuous Dynamic Graph Neural Network for Graph Classification
  Jie Liu (Southwest University); Jiamou Liu (The University of Auckland); Kaiqi Zhao (University of Auckland); Yanni Tang (University of Auckland); Wu Chen (College of Computer & Information Science, Southwest University)
- GraphHI: Boosting Graph Neural Networks for Large-Scale Graphs
  Hao Feng (Tsinghua University); Chaokun Wang (Tsinghua University); Ziyang Liu (Tsinghua University); Yunkai Lou (Tsinghua University); Zhenyu Liu (Tsinghua University); xiaokun zhu (JD.COM); Yongjun Bao (JD.com); Weipeng Yan (JD.com)
- DiscoGNN: A Sample-Efficient Framework for Self-Supervised Graph Representation Learning
  Jun Xia (Westlake University); Shaorong Chen (Westlake University and Zhejiang University); Yue Liu (NUDT); Zhangyang Gao (westlake university); Jiangbin Zheng (Westlake University); Xihong Yang (National University of Defense Technology); Stan Z. Li (Westlake University)
- Incorporating Dynamic Temperature Estimation into Contrastive Learning on Graphs
  Ziyang Liu (Tsinghua University); Chaokun Wang (Tsinghua University); Liqun Yang (Nanhu Academy); Yunkai Lou (Tsinghua University); Hao Feng (Tsinghua University); Cheng Wu (Tsinghua University); Kai Zheng (Kuaishou); Yang Song (Kuaishou Technology)
- Newton sketches: Estimating Node Intimacy in Dynamic Graphs Using Newton's Law of Cooling
  Qizhi Chen (Peking University); Ke Wang (Yale University); Aoran Li (Peking University); Yuhan Wu (Peking University); Tong Yang (Peking University); Bin Cui (Peking University)
- Counting Butterflies in Fully Dynamic Bipartite Graph Streams
  Serafeim Papadias (TU Berlin); Varun Pandey (TU Berlin); Zoi Kaoudi (IT University of Copenhagen); Jorge Arnulfo Quiané Ruiz (IT University of Copenhagen); Volker Markl (Technische Universität Berlin)
- BIM: Improving Graph Neural Networks with Balanced Influence Maximization
  Wentao Zhang (Peking University); Xinyi Gao (The University of Queensland); Ling Yang (Peking University); Meng Cao (Apple); Ping H (Apple); Jiulong Shan (Apple Inc.); Hongzhi Yin (The University of Queensland); Bin Cui (Peking University)

Modern Hardware and In-Memory Database Systems III - [In Theater 11, Chair: Nikos Ntarmos]

- UltraPrecise: A GPU-Based Framework for Arbitrary-Precision Arithmetic in Database Systems
  Xin Li (Shandong University); Mengbai Xiao (Shandong University); Dongxiao Yu (Shandong University); Rubao Lee (Freelance); Xiaodong Zhang (Ohio State U.)
- Exploiting Persistent CPU Cache for Scalable Persistent Hash Index
  Bowen Zhang (Shanghai Jiao Tong University); Shengan Zheng (MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University); Liangxu Nie (Shanghai Jiao Tong University); Zhenlin Qi (Shanghai Jiao Tong University); Linpeng Huang (Shanghai Jiao Tong University); Hong Mei (Shanghai Jiao Tong University)
- LTPG: Large-Batch Transaction Processing on GPUs with Deterministic Concurrency Control
  Jianpeng Wei (Northeastern University); Yu Gu (Northeastern University); Tianyi Li (Aalborg University); Jianzhong Qi (The University of Melbourne); Chuanwen Li (Northeastern University); Yanfeng Zhang (Northeastern University); Christian S. Jensen (Aalborg University); Ge Yu (Northeastern University)
- Why Files If You Have a DBMS?
  Lam-Duy Nguyen (Technische Universität München); Viktor Leis (Technische Universität München)
- STEM: Streaming-based FPGA Acceleration for Large-Scale Compactions in LSM KV
  Dongdong TANG (City University of Hong Kong); Weilan Wang (City University of Hongkong); Jinghuan YU (City University of Hong Kong); Tei-Wei Kuo (National Taiwan University); Chun Jason XUE (MBZUAI)
- Improving the Relationship between B+-Tree and Memory Allocator for Persistent Memory
  Wei Yan (Xi'an Jiaotong University); Xingjun Zhang (Xi'an Jiaotong University)
- Accelerating Aggregation using Real Processing-in-Memory System
  Muhammad Attahir Jibril (TU Ilmenau); Hani Al-Sayeh (TU Ilmenau); Kai-Uwe Sattler (TU Ilmenau)

Graphs, Networks, and Semistructured Data XI - [In Theater 10, Chair: Zhenhua Huang]

- SES: Bridging the Gap Between Explainability and Prediction of Graph Neural Networks
  zhenhua huang (Anhui University); Li Kunhao (Anhui University); Shaojie Wang (Anhui University); Zhaohong Jia (Anhui University); Wentao Zhu (Amazon); Sharad Mehrotra (U.C. Irvine)
- Efficient Cross-layer Community Search in Large Multilayer Graphs
  Longxu SUN (Hong Kong Baptist University); Xin Huang (Hong Kong Baptist University); Zheng WU (Hong Kong Baptist University); Jianliang Xu (Hong Kong Baptist University)
- Large Subgraph Matching: A Comprehensive and Efficient Approach for Heterogeneous Graphs
  Hongtai Cao (University of Illinois at Urbana-Champaign); Qihao Wang (University of Illinois at Urbana Champaign); Xiaodong Li (The University of Hong Kong); Mohammad Matin Najafi (The University of Hong Kong); Kevin Chen-Chuan Chang (University of Illinois at Urbana-Champaign); Reynold Cheng ("The University of Hong Kong, China")
- Adaptive Hypergraph Network for Trust Prediction
  Rongwei Xu (Macquarie University); Guanfeng Liu (Macquarie University); Yan Wang (Macquarie University, Austrilia); Xuyun Zhang (Macquarie University); Kai Zheng (University of Electronic Science and Technology of China); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- Bottom-up k-Vertex Connected Component Enumeration by Multiple Extension
  Haoyu Liu (Renmin University of China); Yongcai Wang (Renmin University of China); XIAOJIA XU (Renmin University of China); Deying Li (School of information, Renmin University of China)
- Wings: Efficient Online Multiple Graph Pattern Matching
  Guanxian Jiang (CUHK); Yunjian Zhao (CUHK); Yichao Li (CUHK); Zhi Liu (CUHK); Tatiana Jin (CUHK); Wanying Zheng (CUHK); Boyang Li (CUHK); James Cheng (CUHK)
- SGCL: Semantic-aware Graph Contrastive Learning with Lipschitz Graph Augmentation
  Jinhao Cui (Harbin Institute of Technology, Shen Zhen); Heyan Chai (HarBin Institute of technology, ShenZhen); Xu Yang (Harbin Institute of Technology, Shenzhen); Ye Ding (Dongguan University of Technology); Binxing Fang (Chinese Academy of Engineering); Qing Liao (Harbin Institute of Technology (Shenzhen))

Graphs, Networks, and Semistructured Data XII - [In Theater 9, Chair: Davide Mottin]

- Accelerating Scalable Graph Neural Network Inference with Node-Adaptive Propagation
  Xinyi Gao (The University of Queensland); Wentao Zhang (Peking University); Junliang Yu (The University of Queesland); Yingxia Shao (BUPT); Quoc Viet Hung Nguyen (Griffith University); Bin Cui (Peking University); Hongzhi Yin (The University of Queensland)
- Graph Condensation for Inductive Node Representation Learning
  Xinyi Gao (The University of Queensland); Tong Chen (The University of Queensland); Yilong Zang (Wuhan University); Wentao Zhang (Peking University); Quoc Viet Hung Nguyen (Griffith University); Kai Zheng (University of Electronic Science and Technology of China); Hongzhi Yin (The University of Queensland)
- Graphix: “One User’s JSON is Another User’s Graph”
  Glenn J Galvizo (University of California, Irvine); Michael Carey (UC Irvine)
- CSM-TopK: Continuous Subgraph Matching with TopK Density Constraints
  Chuchu Gao (Hunan University); Youhuan Li (Hunan University); Zhibang Yang (Hunan University); Xu Zhou (Hunan university)
- Efficient Maximal Temporal Plex Enumeration
  Yanping Wu (University of Technology Sydney); Renjie Sun (East China Normal University); Xiaoyang Wang (University of New South Wales); Ying Zhang (University of Technology Sydney); Lu Qin (UTS); Wenjie Zhang (University of New South Wales); Xuemin Lin (Shanghai Jiaotong University)
- Denoising High-Order Graph Clustering
  Yonghao Chen (Guangdong university of technology); Ruibing Chen (Guangdong University of Technology); Qiaoyun Li (Guangdong University of Technology); Xiaozhao Fang (Guangdong University of Technology); JiaXing Li (Institute of Textiles and Clothing, The Hong Kong Polytechnic University); Waikeung Wong (School of Fashion and Textiles, The Hong Kong Polytechnic University)
- A Revisit to Graph Neighborhood Cardinality Estimation
  Pinghui Wang (Xi'an Jiaotong University); Yuanming Zhang (Xi'an Jiaotong University); Kuankuan Cheng (Xi'an Jiaotong University); Junzhou Zhao (Xi'an Jiaotong University)

Data Stream Systems and Edge Computing II - [In Theater 4, Chair: Yunfan Kang]

- A Predictive Profiling and Performance Modeling Approach for Distributed Stream Processing in Edge
  Hasan Geren (RMIT University); Nasrin Sohrabi (Deakin University); Zahir Tari (RMIT University); Nour Moustafa (UNSW Canberra at ADFA)
- Joint Mobile Edge Caching and Pricing: A Mean-Field Game Approach
  Yin Xu (University of Science and Technology of China); Xichong Zhang (University of Science and Technology of China); Mingjun Xiao (University of Science and Technology of China); Jie Wu (Temple University); An Liu (Soochow University); Sheng Zhang (Nanjing University)
- Online Container Caching with Late-Warm for IoT Data Processing
  Guopeng Li (University of Science and Technology of China); Haisheng Tan (University of Science and Technology of China); Xuan Zhang (University of Science and Technology of China); Chi Zhang (University of Science and Technology of China); Ruiting Zhou (Southeast University); Zhenhua Han (Microsoft Research Asia); Guoliang Chen (University of Science and Technology of China)
- COUPLE: Orchestrating Video Analytics on Heterogeneous Mobile Processors
  Hao Bao (Sun Yat-sen University); Zhi Zhou (Sun yat-sen University); Fei Xu (East China Normal University); Xu Chen (Sun Yat-sen University)
- Multiple Continuous Top-K Queries Over Data Stream
  Rui Zhu (Shenyang Aerospace University); Yujin Jia (Shenyang Aerospace University); Xiaochun Yang (Northeastern University); Baihua Zheng (Singapore Management University); Bin Wang (Northeastern University); Chuanyu Zong (Shenyang Aerospace University)
- CodingSketch: A Hierarchical Sketch with Efficient Encoding and Recursive Decoding
  Qizhi Chen (Peking University); Yisen Hong (Peking University); Yuhan Wu (Peking University); Tong Yang (Peking University); Bin Cui (Peking University)
- Everything Everyway All at Once - Time Traveling Debugging for Stream Processing Applications
  Timo Räth (TU Ilmenau); Marius Schlegel (TU Ilmenau); Kai-Uwe Sattler (TU Ilmenau)

Data Mining and Knowledge Discovery XII - [In Theater 3, Chair: Quanyu Dai]

- A Two Phase Recall-and-Select Framework for Fast Model Selection
  Jianwei Cui (Xiaomi AI Lab.); WenHang Shi (Renmin University of China); Honglin Tao (Renmin University of China); WEI LU (Renmin University of China); Xiaoyong Du (Renmin University of China)
- BTS: Load-Balanced Distributed Union-Find for Finding Connected Components with Balanced Tree Structures
  Chaeeun Kim (University of California, Santa Cruz); Changhun Han (Kookmin University); Ha-Myung Park (Kookmin University)
- Interpretable Knowledge Tracing via Response Influence-based Counterfactual Reasoning
  Jiajun Cui (East China Normal University); Minghe Yu (Northeastern University); Bo Jiang (East China Normal University); Aimin Zhou (East China Normal University); Jianyong Wang (Tsinghua University); Wei Zhang (East China Normal University)
- Stable Heterogeneous Treatment Effect Estimation across Out-of-Distribution Populations
  Yuling Zhang (Tsinghua University); Anpeng Wu (Zhejiang University); Kun Kuang (Zhejiang University); Liang Du (Tencent); Zixun Sun (tencent); Zhi Wang (Tsinghua University)
- Towards Cross-Domain Continual Learning
  Marcus de Carvalho (Nanyang Technological University); mahardhika pratama (University of South Australia); Jie Zhang (Nanyang Technological University); Haoyan Chua (Nanyang Technological University); Edward KY Yapp (SIMTech)
- DROPP: Structure-aware PCA for Ordered Data - A General Method and its Applications in Climate Research and Molecular Dynamics
  Anna Beer (University of Vienna); Olivér Palotás (Aarhus University); Andrea Maldonado (LMU Munich); Andrew Draganov (Aarhus University); Ira Assent (Aarhus University)

Database Security and Privacy II - [In Theater 2, Chair: Sharad Mehrotra]

- SecMdp: Towards Privacy-Preserving Multimodal Deep Learning in End-Edge-Cloud
  Zhao Bai (Beijing Normal University); Mingyue Wang (Harbin Institute of Technology, Shenzhen); Fangda Guo (Institute of Computing Technology, Chinese Academy of Sciences); Yu Guo (Beijing Normal University); Chengjun Cai (City University of Hong Kong); Rongfang Bie (Beijing Normal University); Xiaohua Jia (City University of Hong Kong)
- CARGO: Crypto-Assisted Differentially Private Triangle Counting without Trusted Servers
  Shang Liu (Kyoto University); Yang Cao (Hokkaido University); Takao Murakami (The Institute of Statistical Mathematics (ISM)); Jinfei Liu (Zhejiang University); Masatoshi Yoshikawa (Osaka Seikei University)
- Real-Time Trajectory Synthesis with Local Differential Privacy
  Yujia Hu (Zhejiang University); Yuntao Du (Purdue University); Zhikun Zhang (Stanford University); Ziquan Fang (Zhejiang University); Lu Chen (Zhejiang University); Kai Zheng (University of Electronic Science and Technology of China); Yunjun Gao (Zhejiang University)
- Privacy-preserving Traffic Flow Release with Consistency Constraints
  Xiaoting Zhu (Sun Yat-sen University); Libin Zheng (Sun Yat-sen University); Chen Zhang (The Hong Kong Polytechnic University); Peng Cheng (East China Normal University); Rui Meng (BNU-HKBU United International College); Lei Chen (Hong Kong University of Science and Technology); Xuemin Lin (Shanghai Jiaotong University); Jian Yin (Sun Yat-Sen University)
- Unraveling Privacy Risks of Individual Fairness in Graph Neural Networks
  He Zhang (Monash University); Xingliang Yuan (Monash University); Shirui Pan (Griffith University)
- Sketches-based join size estimation under local differential privacy
  Meifan Zhang (Guangzhou University); Xin Liu (Guangzhou University); Lihua Yin (Guangzhou University)
- PrivShape: Extracting Shapes in Time Series under User-Level Local Differential Privacy
  Yulian MAO (SUSTech & PolyU); Qingqing Ye (Hong Kong Polytechnic University); Haibo Hu (Hong Kong Polytechnic University); Qi Wang (Southern University of Science and Technology); Kai Huang (Macau University of Science and Technology)

Tutorial on Quantum data management: from theory to opportunities [In Theater 1]

Abstract:
Quantum computing has emerged as a transformative tool for future data management. Classical problems in database domains, including query optimization, data integration, and transaction management, have recently been addressed using techniques from quantum computing. This tutorial aims to establish the theoretical foundation essential for enhancing methodologies and practical implementations in this line of research.
Moreover, this tutorial takes a forward-looking approach by delving into recent strides in quantum internet technologies and the nonlocality theory. We aim to shed light on the uncharted territory of future data systems tailored for the quantum internet.

Presenters:

Rihan Hai is an assistant professor at TU Delft, Netherlands. Her research focuses on data management for machine learning, federated learning, and quantum data management. She has served as a PC member of VLDB, and ICDE, and a journal reviewer for TKDE, VLDBJ, SIGMOD Record, JMLR and TPDS.

Shih-Han Hung is a postdoc at Academia Sinica. His research aims to better understand the power and the limit of quantum computers. Previously, he was a postdoc at the University of Texas at Austin. He received his Ph.D. from the University of Maryland.

Sebastian Feld is an assistant professor at Quantum & Computer Engineering department of TU Delft, Netherlands. He and his group are working on Quantum Machine Learning. Before, he was head of Quantum Applications and Research Laboratory (QAR-Lab) at LMU Munich.

Poster Session - [In Speys]

Posters of the 13:15 and 15:00 Session

Poster Session - [In B2B]

Posters of the 13:15 and 15:00 Session

Friday May 17th, 2024 8:30-10:00

Text, Semi-Structured Data, IR, Image, and Multimedia databases II - [In Theater 12, Chair: TBD]

- HIT: Solving Partial Index Tracking via Hierarchical Reinforcement Learning
  Zetao Zheng (University of Electronic Science and Technology of China); Jie Shao (University of Electronic Science and Technology of China); Feiyu Chen (University of Electronic Science and Technology of China); Anjie Zhu (University of Electronic Science and Technology of China); Shilong Deng (University of Electronic Science and Technology of China); Heng Tao Shen (University of Electronic Science and Technology of China (UESTC))
- FieldSwap: Data Augmentation for Effective Form-Like Document Extraction
  Jing Xie (Google); James B Wendt (Google); Yichao Zhou (Google); Seth Ebner (Johns Hopkins University); Sandeep Tata (Google)
- LT^{2}R: Learning to Online Learning to Rank for Web Search
  Xiaokai Chu (Institute of Computing Technology, Chinese Academy of Sciences); Lixin Zou (Wuhan University); Chenliang Li (Wuhan University); Jiashu Zhao (Wilfrid Laurier University); Changying Hao (Baidu Inc.); Shuaiqiang Wang (Baidu Inc.); Dawei Yin (Baidu)
- MUST: An Effective and Scalable Framework for Multimodal Search of Target Modality
  Mengzhao Wang (Zhejiang University); Xiangyu Ke (Zhejiang University, China); Xiaoliang Xu (Hangzhou Dianzi University); Lu Chen (Zhejiang University); Yunjun Gao (Zhejiang University); Pinpin Huang (Hangzhou Dianzi University); Runkai Zhu (Hangzhou Dianzi University)
- Online Anomaly Detection over Live Social Video Streaming
  Chengkun He (RMIT University); Xiangmin Zhou (RMIT University); Chen Wang (DATA61, CSIRO); Iqbal Gondal (RMIT University); Jie Shao (University of Electronic Science and Technology of China); Xun Yi (RMIT University)

Query Processing, Indexing, and Optimization IV - [In Theater 11, Chair: TBD]

- CLIMBER: Pivot-Based Approximate Similarity Search over Big Data Series
  Liang Zhang (Oracle); Mohamed Y. Eltabakh (Qatar Computing Research Institute (QCRI)); Elke A Rundensteiner (Worcester Polytechnic Institute); Khalid Alnuaim (Worcester Polytechnic Institute)
- Hill-Cache: Adaptive Integration of Recency and Frequency in Caching with Hill-Climbing
  Yunfan Li (East China Normal University); Huiqi Hu (East China Normal University); Chaojing Lei (Zhejiang University); Xuan Zhou (East China Normal University); Weining Qian (East China Normal University)
- Efficient Approximate Maximum Inner Product Search over Sparse Vectors
  Xi Zhao (HKUST); Zhonghan Chen (HKUST); Kai Huang (Macau University of Science and Technology); Ruiyuan Zhang (The Hong Kong University of Science and Technology); Bolong Zheng (Huazhong University of Science and Technology); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- Riveter: Adaptive Query Suspension and Resumption Framework for Cloud Native Databases
  Rui Liu (University of Chicago); Aaron J Elmore (University of Chicago); Michael Franklin (University of Chicago); Sanjay Krishnan (UChicago)
- Top-L Most Influential Community Detection Over Social Networks
  Nan Zhang (East China Normal University); Yutong Ye (East China Normal University); Xiang Lian (Kent State University); Mingsong Chen (East China Normal University)

Graphs, Networks, and Semistructured Data XIII - [In Theater 10, Chair: Yuchen Li]

- Attributed Network Embedding in Streaming Style
  Anbiao Wu (Northeastern University); Ye Yuan (Beijing Institute of Technology); Changsheng Li (Beijing Institute of Technology); Yuliang Ma (Northeastern University); Hao Zhang (Huawei)
- Faster Depth-First Subgraph Matching on GPUs
  Lyuheng Yuan (Indiana University Bloomington); Da Yan (Indiana University Bloomington); Jiao Han (George Washington University); Akhlaque Ahmad (TED University); Yang Zhou (Auburn University); Zhe Jiang (University of Florida)
- G2-AIMD: A Memory-Efficient Subgraph-Centric Framework for Efficient Subgraph Search on GPUs
  Lyuheng Yuan (Indiana University Bloomington); Akhlaque Ahmad (TED University); Da Yan (Indiana University Bloomington); Jiao Han (George Washington University); Saugat Adhikari (Indiana University Bloomington); Xiaodong Yu (Stevens Institute of Technology); Yang Zhou (Auburn University)
- Fine-grained Anomaly Detection on Dynamic Graphs via Attention Alignment
  Dong Chen (National University of Defense Technology); Xiang Zhao (National University of Defense Technology); Weidong Xiao (National University of Defense Technology)
- Accelerating Biclique Counting on GPU
  Linshan Qiu (Zhejiang University); Zhonggen Li (Zhejiang University); Xiangyu Ke (Zhejiang University, China); Lu Chen (Zhejiang University); Yunjun Gao (Zhejiang University)

Uncertain, Probabilistic, and Approximate Databases - [In Theater 9, Chair: Saeid Hosseini]

- Computing All Restricted Skyline Probabilities on Uncertain Datasets
  Xiangyu Gao (Harbin Institute of Technology); Jianzhong Li (Harbin Institute of Technology); Dongjing Miao (Harbin Institute of Technology)
- M4: A Framework for Per-Flow Quantile Estimation
  Siyuan Dong (Peking University); Zhuochen Fan (Peking University); Tianyu Bai (Peking University); Tong Yang (Peking University); Hanyu Xue (Peking University); Peiqing Chen (Peking University); Yuhan Wu (Peking University)
- DISCO: A Dynamically Configurable Sketch Framework in Skewed Data Streams
  Jiaqian Liu (Nanjing University); Ran Ben Basat (University College London); Louis de Wardt (University College London); Haipeng Dai (Nanjing University); Guihai Chen (Nanjing University)
- BitMatcher: Bit-level Counter Adjustment for Sketches
  Qilong Shi (Peking University); Chengjun Jia (Tsinghua University); Wenjun Li (Harvard University); Zaoxing Liu (University of Maryland); Tong Yang (Peking University); Jianan Ji (Peking University); Gaogang Xie (Computer Network Information Center, Chinese Academy of Sciences); Weizhe Zhang (Pengcheng Laboratory); Minlan Yu (Harvard)
- Space-efficient indexes for uncertain strings
  Estéban Gabory (CWI); Chang Liu (Zhejiang University); Grigorios Loukides (King's College London); Solon Pissis (CWI); Wiktor Zuba (CWI)

Benchmarking, Performance Modeling, Tuning, and Testing [In Theater 4, Chair: George Christodoulou]

- Mirage: Generating Enormous Databases for Complex Workloads
  Qingshuai Wang (East China Normal University); Hao Li (East China Normal University); Zirui Hu (East China Normal University); Rong Zhang (East China Normal University); Chengcheng Yang (East China Normal University); Peng Cai (East China Normal University); Xuan Zhou (East China Normal University); Aoying Zhou (East China Normal University)
- Joint Directory, File and IO Trace Feature Extraction and Feature-based Trace Regeneration for Enterprise Storage Systems
  Kecheng HUANG (The Chinese University of Hong Kong); Xijun Li (Huawei Noah's Ark Lab); Mingxuan Yuan (Huawei); Ji Zhang (Huawei Technologies Co., Ltd); Zili Shao (The Chinese University of Hong Kong)
- Robust Auto-Scaling with Probabilistic Workload Forecasting for Cloud Databases
  Haitian Hang (Zhejiang University); Xiu Tang (Zhejiang University); Jianling Sun (Zhejiang University); Lingfeng Bao (Zhejiang University); David Lo (Singapore Management University); Haoye Wang (Hangzhou City University)
- CheckMate: Evaluating Checkpointing Protocols for Streaming Dataflows
  George Siachamis (TU Delft); Kyriakos Psarakis (TU Delft); Marios Fragkoulis (Delivery Hero); Arie Van Deursen (Delft University of Technology); Paris Carbone (KTH Royal Institute of Technology); Asterios Katsifodimos (TU Delft)
- BenchTemp: A General Benchmark for Evaluating Temporal Graph Neural Networks
  Qiang Huang (Wuhan University); xin wang (Wuhan university); Susie Xi Rao (ETH); Zhichao Han (Ebay); Zitao Zhang (eBay); Yongjun He (ETH Zürich); Quanqing Xu (OceanBase, Ant Group); Yang Zhao (Ebay); Zhigao Zheng (Wuhan University); Jiawei Jiang (Wuhan University)

Workflows, Scientific Data Management - [In Theater 3, Chair: TBD]

- Enhancing LSM-tree Key-Value Stores for Read-Modify-Writes via Key-Delta Separation
  Jinhong Li (The Chinese University of Hong Kong); Yanjing Ren (The Chinese University of Hong Kong); Shujie Han (Peking University); Patrick P. C. Lee (The Chinese University of Hong Kong)
- TMan: A High-Performance Trajectory Data Management System Based on Key-value Stores
  Huajun He (School of Computing and Artificial Intelligence, Southwest Jiaotong University); Xu ZiHang (Southwest Jiaotong University); Ruiyuan Li (Chongqing University); Jie Bao (JD Intelligent Cities Research); Tianrui Li (School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, 611756, China); Yu Zheng (JD)
- Kondo: Efficient Provenance-driven Data Debloating
  Aniket Modi (IIT Delhi); Rohan Tikmany (DePaul University); Tanu Malik (DePaul University); Raghavan Komondoor (Indian Institute of Science); Ashish Gehani (SRI); Deepak D'Souza (Indian Institute of Science, Bangalore)
- Preserving Topological Feature with Sign-of-Determinant Predicates in Lossy Compression: A Case Study of Vector Field Critical Points
  Mingze Xia (University of Kentucky); Sheng Di (Argonne National Laboratory, Lemont, IL); Franck Cappello (Argonne National Laboratory, Lemont, IL); Pu Jiao (University of Kentucky); Kai Zhao (Florida State University); Jinyang Liu (University of California, Riverside); Xuan Wu (University of Kentucky); Xin Liang (University of Kentucky); Hanqi Guo (The Ohio State University)
- FreqyWM: Frequency Watermarking for the New Data Economy
  Devriş İşler (IMDEA Networks & UC3M); Elisa Cabana (CUNEF Universidad); Alvaro Garcia-Recuero (Fundación para el Desarrollo y la Innovación Tecnológica (FUNDITEC)); Georgia Koutrika (ATHENA Research Center); Nikolaos Laoutaris (IMDEA Networks)

Data Mining and Knowledge Discovery XIV - [In Theater 2, Chair: Rihan Hai]

- Scalable Overspeed Item Detection in Streams
  Yuhan Wu (Peking University); Hanbo Wu (Peking University); Chengjun Jia (Tsinghua University); Bo Peng (Peking University); Ziyun Zhang (Peking University); Tong Yang (Peking University); Peiqing Chen (Peking University); Kaicheng Yang (Peking University)
- GradGCL: Gradient Graph Contrastive Learning
  Ran Li (Hong Kong University of Science and Technology); Shimin Di (The Hong Kong University of Science and Technology); Lei Chen (Hong Kong University of Science and Technology); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- ST-ABC: Spatio-Temporal Attention-Based Convolutional Network for Multi-Scale Lane-Level Traffic Prediction
  Shuhao Li (FudanUniversity); Yue Cui (The Hong Kong University of Science and Technology); Libin Li (Guangzhou University); Weidong Yang (Fudan University); Fan Zhang (Guangzhou University); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- CPDG: A Contrastive Pre-Training Method for Dynamic Graph Neural Networks
  Yuanchen Bei (Zhejiang University); Hao Xu (meituan); Sheng Zhou (Zhejiang University); Huixuan Chi (Institute of Computing Technology, Chinese Academy of Sciences); Haishuai Wang (Zhejiang University); Mengdi Zhang (Meituan); Zhao Li (Zhejiang University); Jiajun Bu (Zhejiang University)
- Graph Anomaly Detection at Group Level: A Topology Pattern Enhanced Unsupervised Approach
  Xing AI (The Hong Kong Polytechnic University); Jialong Zhou (The Hong Kong Polytechnic University); Yulin Zhu (Hong Kong Polytechnic University); Gaolei Li (Shanghai Jiaotong University); Tomasz Michalak (University of Warsaw & Ideas NCBiR); Xiapu Luo (Hong Kong Polytechnic University); Kai Zhou (Hong Kong Polytechnic University)

Tutorial on An Interactive Dive into Time-Series Anomaly Detection [In Theater 1]

Abstract:
Anomaly detection is an important problem in data analytics with applications in many domains. In recent years, there has been an increasing interest in anomaly detection tasks applied to time series. In this tutorial, we take a holistic view of anomaly detection in time series, starting from the core definitions and taxonomies related to time series and anomaly types, to an extensive description of the anomaly detection methods proposed by different communities in the literature. We explore the literature and the proposed methods by demonstrating systems that help users understand the core computational steps of some methods and navigate benchmark results. Finally, we describe the problem of model selection for anomaly detection and discuss recent experimental results.

Presenters:

Paul Boniol is a researcher at Inria, member of the VALDA project-team. Previously, he worked at ENS Paris-Saclay (Centre Borelli), Universit´e Paris Cit´e, EDF Research lab, and Ecole Polytechnique (LIX). His research interests lie between data analytics, machine learning, and time-series analysis. His Ph.D. dissertation focused on subsequence anomaly detection and time-series classification. His work has been published in the top data management and analytics venues.

John Paparrizos is an assistant professor at The Ohio State University, leading The DATUM Lab. His research focuses on adaptive solutions for data-intensive and machine-learning applications. His doctoral work was recognized at the 2019 ACM SIGKDD Doctoral Dissertation Award competition. He has also received the inaugural ACM SIGMOD Research Highlight Award, a NetApp Faculty Award, and the 2023 IEEE TCDE Rising Star Award. His ideas have been widely adopted across scientific areas, Fortune 100-500 companies (e.g., Exelon and Nokia), and organizations such as ESA.

Themis Palpanas is an elected Senior Member of the French University Insitute (IUF), and Distinguished Professor of computer science at the University of Paris (France). He is the author of 14 patents, has received 3 best paper awards and the IBM SUR award, has been Program Chair for VLDB 2025 and IEEE BigData 2023, General Chair for VLDB 2013, and has served Editor in Chief for BDR. He has been working in the fields of Data Series Management and Analytics for more than 15 years, and has developed several of the state of the art techniques. He has delivered 19 tutorials in top conferences.

Poster Session - [In Speys]

Posters of the 16:20 Session

Poster Session - [In B2B]

Posters of the 16:20 Session

Friday May 17th, 2024 10:00-10:30

Coffee Break - [In Speys]

Friday May 17th, 2024 10:30-12:45

Future Technologies 2 - [In Theater 12, Chair: C. Mohan]

- Secure Normal Form: Mediation Among Cross Cryptographic Leakages in Encrypted Databases.
  by Shufan Zhang (University of Waterloo), Xi He (University of Waterloo), Ashish Kundu (Cisco Research), Sharad Mehrotra (U.C. Irvine), Shantanu Sharma (New Jersey Institute of Technology)
- Reactive Knowledge Management.
  by Stefano Ceri (Politecnico di Milano), Anna Bernasconi (Politecnico di Milano), Alessia Gagliardi (Politecnico di Milano)
- LakeHarbor: Making Structures First-Class Citizens in Data Lakes.
  by Hiroyuki Yamada (The University of Tokyo), Masaru Kitsuregawa (University of Tokyo, Japan), Kazuo Goda (Institute of Industrial Science, The University of Tokyo)
- A CXL-Powered Database System: Opportunities and Challenges
  by Yunyan Guo (Tsinghua University), Guoliang Li (Tsinghua University)
- Bifrost: A Future Graph Database Runtime.
  by James Clarkson (Neo4j), Georgios Theodorakis (Neo4j), Jim Webber (Neo4J)
- V2V: Efficiently Synthesizing Video Results for Video Queries.
  by Dominik Winecki (The Ohio State University), Arnab Nandi (The Ohio State University)
- Higher-Order SQL Lambda Functions.
  by Maximilian Schüle (University of Bamberg), Jakob Hornung (Universität Bamberg)

Friday May 17th, 2024 12:45-14:00

Lunch - [In Speys]

Friday May 17th, 2024 14:00-15:50

AI for Databases IV - [In Theater 12, Chair: TBD]

- Dynamic Data Layout Optimization with Worst-case Guarantees
  Kexin Rong (Georgia Institute of Technology); Paul Liu (Stanford University); Sarah Ashok Sonje (Georgia Institute of Technology); Moses Charikar (Stanford University, California)
- QCFE: An Efficient Feature Engineering for Query Cost Estimation.
  Yu Yan (Harbin Institute of Technology); Hongzhi Wang (Harbin Institute of Technology); Huang Junfang (Harbin Institute of Technology); Dake Zhong (Harbin Institute of Technology); Tao Yu (Harbin Institute of Technology); Kaixin Zhang (Harbin Institute of Technology); Man Yang (Harbin Institute of Technology); Tianqing Wang (Huawei)
- Chameleon: Towards Update-Efficient Learned Indexing for Locally Skewed Data
  Na Guo (Northeastern University); Yaqi Wang (Shenyang Aerospace University); Wenli L Sun (Shenyang Aerospace University); Yu Gu (Northeastern University); Jianzhong Qi (The University of Melbourne); Zhenghao Liu (Northeastern University); Xia Xiufeng (Shenyang Aerospace University); Ge Yu (Northeastern University)
- FOSS: A Self-Learned Doctor for Query Optimizer
  Kai Zhong (Renmin Uninversity of China); Luming Sun (Renmin University of China); Tao Ji (Renmin University of China); Cuiping Li (Renmin University of China); Hong Chen (" Renmin University, China")
- MFIX: An Efficient and Reliable Index Advisor via Multi-Fidelity Bayesian Optimization
  Zhuo Chang (Peking University); Xinyi Zhang (Peking University); Yang Li (Tencent Inc.); Xupeng Miao (Carnegie Mellon University); Yanzhao Qin (Peking University); Bin Cui (Peking University)
- VDTuner: Automated Performance Tuning for Vector Data Management Systems
  Tiannuo Yang (Nankai University); Wangqi Peng (Nankai University); Yusen Li (Nankai University); Wang Gang (Nankai Univerisity); Liu Xiaoguang (Nankai Univerisity); Wen Hu (Ant Group); Jianguo Li (Ant Group)

Query Processing, Indexing, and Optimization V - [In Theater 11, Chair: Patrick P. C. Lee]

- Fast Query Answering by Labeling Index on Uncertain Graphs
  Zeyu Wang (Zhejiang University); Qihao Shi (Zhejiang University); Jiawei Chen (Zhejiang University); Can Wang (Zhejiang University); Mingli Song (Zhejiang University); Xinyu Wang (Zhejiang University)
- Scavenger: Better Space-Time Trade-Offs for Key-Value Separated LSM-trees
  Jianshun Zhang (Huazhong University of Science and Technology); Fang Wang (Huazhong University of Science and Technology); Sheng Qiu (ByteDance); Yi Wang (bytedance); Jiaxin Ou (Bytedance); Junxun Huang (Huazhong University of Science and Technology); Baoquan Li (Huazhong University of Science and Technology); Peng Fang (Huazhong University of Science and Technology); Dan Feng (Huazhong University of Science and Technology)
- A Spatio-Temporal Series Data Model with Efficient Indexing and Layout for Cloud-Based Trajectory Data Management
  Yang Guo (The Chinese University of Hong Kong); Zhiqi Wang (The Chinese University of Hong Kong); Jin XUE (The Chinese University of HK); Zili Shao (The Chinese University of Hong Kong)
- Reverse Regret Query
  Weicheng Wang (Hong Kong University of Science and Technology); Raymond Chi-Wing Wong (Hong Kong University of Science and Technology); H. V. Jagadish (University of Michigan); Min Xie (Shenzhen Institute of Computing Sciences)
- Resistance Eccentricity in Graphs: Distribution, Computation and Optimization
  Zenan Lu (Fudan University); Xiaotian Zhou (Fudan University); Ahad N. Zehmakan (The Australian National University); Zhongzhi Zhang (Fudan University)
- BushStore: Efficient B+Tree Group Indexing for LSM-Tree in Non-Volatile Memory
  Zhenghao Wang (Zhejiang University); Lidan Shou (Zhejiang University); Ke Chen (Zhejiang University); Xuan Zhou (East China Normal University)

Graphs, Networks, and Semistructured Data XIV - [In Theater 10, Chair: Ziyi Ma]

- GPU-Accelerated Batch-Dynamic Subgraph Matching
  Linshan Qiu (Zhejiang University); Lu Chen (Zhejiang University); Hailiang Jie (Zhejiang University); Xiangyu Ke (Zhejiang University, China); Yunjun Gao (Zhejiang University); Yang Liu (Huawei); Zetao Zhang (Huawei Technologies Co., Ltd.)
- I/O Efficient Max-Truss Computation in Large Static and Dynamic Graphs
  Jiaqi Jiang (Beijing Institute of Technology); Qi Zhang (Beijing Institute of Technology); Ronghua Li (Beijing Institute of Technology); Qiangqiang Dai (Beijing Institute of Technology); Guoren Wang (Beijing Institute of Technology)
- Efficient Multi-query Oriented Continuous Subgraph Matching
  Ziyi Ma (Hebei University of Technology); Jianye Yang (Guangzhou University); Xu Zhou (Hunan university); Guoqing Xiao (Hunan University); Jianhua Wang (Hunan University); Liang Yang (Hebei University of Technology); Kenli Li (Hunan University); Xuemin Lin (Shanghai Jiaotong University)
- Label Constrained Reachability Queries on Time Dependent Graphs
  Yishu Wang (Northeastern University); Jinlong Chu (Northeastern University); Ye Yuan (Beijing Institute of Technology); Yu Gu (Northeastern University); Hangxu Ji (Northeastern University); Hao Zhang (Huawei)
- Time-Constrained Continuous Subgraph Matching Using Temporal Information for Filtering and Backtracking
  Seunghwan Min (Seoul National University); Jihoon Jang (Seoul National University); Kunsoo Park (Seoul National University); Dora Giammarresi (Universita Roma Tor Vergata); Giuseppe F. Italiano (LUISS University); Wook-Shin Han (POSTECH)
- Adaptive Truss Maximization on Large Graphs: A Minimum Cut Approach
  Zitan Sun (HKBU); Xin Huang (Hong Kong Baptist University); Chengzhi Piao (Hong Kong Baptist University); Cheng Long (Nanyang Technological University); Jianliang Xu (Hong Kong Baptist University)

Data Mining and Knowledge Discovery XV - [In Theater 9, Chair: Cristian Consonni]

- Temporal-Frequency Masked Autoencoders for Time Series Anomaly Detection
  Yuchen Fang (University of Electronic Science and Technology of China); jiandong xie (HUAWEI TECHNOLOGIES CO.LTD.); Yan Zhao (Aalborg University); Lu Chen (Zhejiang University); Yunjun Gao (Zhejiang University); Kai Zheng (University of Electronic Science and Technology of China)
- REGER: Reordering Time Series Data for Regression Encoding
  Jinzhao Xiao (Tsinghua University); Wendi He (Tsinghua University); Shaoxu Song (Tsinghua University); Xiangdong Huang (Tsinghua University); Chen Wang (Timecho Limited); Jianmin Wang ("Tsinghua University, China")
- SAGDFN: A Scalable Adaptive Graph Diffusion Forecasting Network for Multivariate Time Series Forecasting
  Yue Jiang (Nanyang Technological University); Xiucheng Li (Harbin Institute of Technology); Yile Chen (Nanyang Technological University); SHUAI LIU (Nanyang Technological University); WEILONG KONG (Nanyang Technological University); Antonis F Lentzakis (NCS); Gao Cong (Nanyang Technological Univesity)
- Knowledge-Enhanced Recommendation with User-Centric Subgraph Network
  Guangyi Liu (Tsinghua University); Quanming Yao (Tsinghua University); Yongqi Zhang (4Paradigm Inc.); Lei Chen (Hong Kong University of Science and Technology)
- MUSE-Net: Disentangling Multi-Periodicity for Traffic Flow Forecasting
  Jianyang Qin (Harbin Institute of Technology (Shenzhen)); Yan Jia (National University of Defense Technology); Yongxin Tong (Beihang University); Heyan Chai (HarBin Institute of technology, ShenZhen); Ye Ding (Dongguan University of Technology); Xuan Wang (Harbin Institute of Technology); Binxing Fang (Chinese Academy of Engineering); Qing Liao (Harbin Institute of Technology (Shenzhen))
- Model Selection with Model Zoo via Graph Learning
  Ziyu Li (Delft University of Technology); Hilco vanderwilk (TU Delft); Danning Zhan (TU Delft); Megha Khosla (TU Delft); Asterios Katsifodimos (TU Delft); Alessandro Bozzon (Delft University of Technology); Rihan Hai (TU Delft)

Spatial Databases and Temporal Databases III - [In Theater 4, Chair: Walid G. Aref]

- Parameterized Decision-making with Multi-modality Perception for Autonomous Driving
  Yuyang Xia (University of Electronic Science and Technology of China); Shuncheng Liu (University of Electronic Science and Technology of China); quanlin yu (University Of Electronic Science And Technology Of China); Liwei Deng (University of Electronic Science and Technology of China); You Zhang (University of Michigan); Han Su (University of Electronic Science and Technology of China); Kai Zheng (University of Electronic Science and Technology of China)
- CausalTAD: Causal Implicit Generative Model for Debiased Online Trajectory Anomaly Detection
  Wenbin Li (Institute of Computing Technology, Chinese Academy of Sciences); Di Yao (Institute of Computing Technology, Chinese Academy of Sciences); Chang Gong (Institute of Computing Technology, Chinese Academy of Sciences); Xiaokai Chu (Institute of Computing Technology, Chinese Academy of Sciences); Quanliang Jing (Institute of Computing Technology, Chinese Academy of Sciences); Xiaolei Zhou (DIDI Global Inc.); Zhang Yuxuan (DiDi Global); Yunxia Fan (DIDI Global Inc.); Jingping Bi (Institute of Computing Technology, Chinese Academy of Sciences)
- Learning to Hash for Trajectory Similarity Computation and Search
  Liwei Deng (University of Electronic Science and Technology of China); Yan Zhao (Aalborg University); Jin Chen (University of Electronic Science and Technology of China(UESTC)); Shuncheng Liu (University of Electronic Science and Technology of China); Yuyang Xia (University of Electronic Science and Technology of China); Kai Zheng (University of Electronic Science and Technology of China)
- Ocean: Online Clustering and Evolution Analysis for Dynamic Streaming Data
  chunhui feng (Soochow University); Junhua Fang (Soochow University); Yue Xia (Soochow University); Pingfu Chao (Soochow University); Pengpeng Zhao (Soochow University); Jiajie Xu (Soochow University); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- SWISP: Distributed Convoy Mining via Sliding Window-based Indexing and Sub-track Partitioning
  Chenxu Wang (Xi'an Jiaotong University); Xin Yang (Xi'an Jiaotong University); Tianyi Li (Aalborg University); Jiaxing Wei (Xi'an Jiaotong University); Pinghui Wang (Xi'an Jiaotong University); Hongzhen Xiang (Xi'an Jiaotong University); Christian S. Jensen (Aalborg University)
- Querying Shortest Path on Large Time-Dependent Road Networks with Shortcuts
  Zengyang Gong (The Hong Kong University of Science and Technology); Yuxiang Zeng (Hong Kong University of Science and Technology); Lei Chen (Hong Kong University of Science and Technology)

Information Integration and Data Quality IV - [In Theater 3, Chair: Karl Aberer]

- Discovering Denial Constraints in Dynamic Datasets
  Eduardo H. M. Pena (UTFPR); Fabio Porto (LNCC); Felix Naumann (Hasso Plattner Institute, University of Potsdam)
- Towards Semantic Consistency: Dirichlet Energy Driven Robust Multi-Modal Entity Alignment
  Yuanyi Wang (Beijing University of Posts and Telecommunications); Haifeng Sun (Beijing university of posts and telecommunications); Jiabo Wang (Beijing university of posts and telecommunications); Jingyu Wang (Beijing University of Posts and Telecommunications); Wei Tang (Huawei); Qi Qi (Beijing University of Posts and Telecommunications); Shaoling Sun (China Mobile (Suzhou) Software Technology Co., Ltd.); Jianxin Liao (beijing university of posts and telecommunications)
- Share: Stackelberg-Nash based Data Markets
  Jinfei Liu (Zhejiang University); Yuran Bi (Zhejiang University); Chen Zhao (Zhejiang University); Junyi Zhao (Zhejiang University); Kui Ren (Zhejiang University); Li Xiong (Emory University)
- Interactive Trimming against Evasive Online Data Manipulation Attacks: A Game-Theoretic Approach
  Yue Fu (The Hong Kong Polytechnic University); Qingqing Ye (Hong Kong Polytechnic University); Rong Du (PolyU); Haibo Hu (Hong Kong Polytechnic University)
- Label Noise Correction for Federated Learning: A Secure, Efficient and Reliable Realization
  Haodi Wang (Beijing Normal University); Tangyu Jiang (Beijing Normal University); Yu Guo (Beijing Normal University); Fangda Guo (Institute of Computing Technology, Chinese Academy of Sciences); Rongfang Bie (Beijing Normal University); Xiaohua Jia (City University of Hong Kong)
- Mitigating Data Scarcity in Supervised Machine Learning through Reinforcement Learning Guided Data Generation
  Chengliang Chai (Beijing Institute of Technology); Kaisen Jin (Beijing Institute of Technology); Nan Tang (HKUST (GZ)); Ju Fan (Renmin University of China); Lianpeng N/A Qiao (Beijing Institute of Technology); Yu-Ping Wang (Beijing Institute of Technology); Yuyu Luo (HKUST (GZ)); Ye Yuan (Beijing Institute of Technology); Guoren Wang (Beijing Institute of Technology)

Data Mining and Knowledge Discovery XVII - [In Theater 2, Chair: TBD]

- Logical Relation Modeling and Mining in Hyperbolic Space for Recommendation
  Yanchao Tan (Fuzhou University); Hang Lv (Fuzhou University); Zihao Zhou (Fuzhou University); Wenzhong Guo (Fuzhou University); Bo Xiong (University of Stuttgart); Weiming Liu (Zhejiang university); Chaochao Chen (Zhejiang University); Shiping Wang (Fuzhou University); Carl Yang (Emory University)
- HeteFedRec: Federated Recommender Systems with Model Heterogeneity
  Wei Yuan (The University of Queensland); Liang Qu (Southern University of Science and Technology); Lizhen Cui (ShanDong University); Yongxin Tong (Beihang University); Xiaofang Zhou (The Hong Kong University of Science and Technology); Hongzhi Yin (The University of Queensland)
- A Compact and Accurate Sketch for Estimating a Large Range of Set Difference Cardinalities
  Peng Jia (Xi'an Jiaotong University); Pinghui Wang (Xi'an Jiaotong University); Rundong Li (Xi'an Jiaotong University); Junzhou Zhao (Xi'an Jiaotong University); Junlan Feng (China Mobile Research); Xidian Wang (China Mobile Group Design Institute); Xiaohong Guan (Xi'an Jiaotong University)
- A Unified Model for Spatio-Temporal Prediction Queries with Arbitrary Modifiable Areal Units
  Liyue Chen (Peking University); Jiangyi Fang (Peking University); Tengfei Liu (China University of Geosciences); Shaosheng Cao (DiDi Global); Leye Wang (Peking University, China)
- Across Images and Graphs for Question Answering
  Zhenyu Wen (Zhejiang University of Technology); Jiaxu Qian (Zhejiang University of Technology); BIN QIAN (Newcastle University); Qin Yuan (Beijing Institute of Technology); Jianbin Qin (Shenzhen Institute of Computing Sciences, Shenzhen University); Ye Yuan (Beijing Institute of Technology); Qi Xuan (Zhejiang University of Technology)

A Comprehensive Tutorial on the over 100 years of Diagrammatic Representations of Logical Statements and Relational Queries [In Theater 1]

Abstract:
Query formulation is increasingly performed by systems that need to guess a user's intent (e.g. via spoken word interfaces). But how can a user know that the computational agent is returning answers to the "right" query? More generally, given that relational queries can become pretty complicated, how can we help users understand existing relational queries, whether human-generated or automatically generated? Now seems the right moment to revisit a topic that predates the birth of the relational model: developing visual metaphors that help users understand relational queries.
This lecture-style tutorial surveys the key visual metaphors developed for diagrammatic representations of logical statements and relational expressions across both the relational database community and the much older diagrammatic reasoning community. We will survey the history and state-of-the art of relationally-complete diagrammatic representations of relational queries, discuss the key visual metaphors developed in over a century of investigating diagrammatic languages, and organize the landscape by mapping their used visual alphabets to the syntax and semantics of Relational Algebra (RA) and Relational Calculus (RC). Tutorial website: https://northeastern-datalab.github.io/diagrammatic-representation-tutorial/

Presenters:

Wolfgang Gatterbauer is an Associate Professor at the Khoury College of Computer Sciences at Northeastern University. His research interests lie in the intersection of theory and practice of data management. He received an NSF Career award and -- with his students and collaborators -- a best paper award at EDBT 2021, best-of-conference mentions for PODS 2021, SIGMOD 2017, WALCOM 2017, and VLDB 2015, and two out of three reproducibility awards for papers published at SIGMOD 2020.

Poster Session - [In Speys]

Posters of the 08:30 and 10:30 Sessions

Poster Session - [In B2B]

Posters of the 08:30 and 10:30 Sessions

Friday May 17th, 2024 15:50-16:20

Coffee Break - [In Speys]

Friday May 17th, 2024 16:20-18:30

AI for Databases V - [In Theater 12, Chair: TBD]

- GLO: Towards Generalized Learned Query Optimization
  Tianyi Chen (Peking University); Jun Gao (Peking University); Yaofeng Tu (ZTE Corporation); Mo Xu (ZTE)
- A Fully On-disk Updatable Learned Index
  Hai Lan (RMIT University); Zhifeng Bao (RMIT University); Shane Culpepper (The University of Queensland); Renata Borovica-Gajic (University of Melbourne); YU DONG (PingCAP Inc.)
- Routing-Guided Learned Product Quantization for Graph-Based Approximate Nearest Neighbor Search
  Qiang Yue (Hangzhou Dianzi University); Xiaoliang Xu (Hangzhou Dianzi University); Yuxiang Wang (Hangzhou Dianzi University); Yikun Tao (Hangzhou Dianzi University); Xuliyuan Luo (Hangzhou Dianzi University)
- Guided SQL-based Data Exploration with User Feedback
  Antonis Mandamadiotis (Athena Research Center); Georgia Koutrika (Athena Research Center); Sihem Amer-Yahia (CNRS)
- ShrinkHPO: Towards Explainable Parallel Hyperparameter Optimization
  TianYu Mu (Harbin Institute of Technology); Hongzhi Wang (Harbin Institute of Technology); Haoyun Tang (National University of Singapore); Xinyue Shao (Harbin Institute of Technology)
- LBSC: A Cost-aware Caching Framework for Cloud Databases
  Zhaoxuan Ji (Beijing Institute of Technology); Zhongle Xie (Zhejiang University); Yuncheng Wu (Renmin University of China); Meihui Zhang (Beijing Institute of Technology)
- DACE: A Database-Agnostic Cost Estimator
  Zibo Liang (University of Electronic Science and Technology of China); Xu Chen (University of Electronic Science and Technology of China); Yuyang Xia (University of Electronic Science and Technology of China); Runfan Ye (University of Electronic Science and Technology of China); HaiTian Chen (University of Electronic Science and Technology of China); jiandong xie (HUAWEI TECHNOLOGIES CO.LTD.); Kai Zheng (University of Electronic Science and Technology of China)

Query Processing, Indexing, and Optimization VI - [In Theater 11, Chair: Maximilian E Schüle]

- FUDJ: Flexible User-Defined Distributed Joins
  Akil Sevim (University of California Riverside); Ahmed Eldawy (University of California, Riverside); Eldon P Carman (Walla Walla University); Michael Carey (UC Irvine); Vassilis J. Tsotras (UC Riverside)
- IVE: Accelerating Emueration-based Subgraph Matching via Exploring Isolated Vertices
  Zite Jiang (ICT, CAS); Shuai Zhang (Institute of Computing Technology, Chinese Academy of Sciences); Xingzhong Hou (Institute of Computing Technology, Chinese Academy of Sciences); Mengting Yuan (Wuhan University); Haihang You (ICT, Chinese Academy of Sciences)
- Approximate Skyline Index for Constrained Shortest Pathfinding with Theoretical Guarantee
  Ziyi Liu (The University of Queensland); Lei Li (The Hong Kong University of Science and Technology (Guang Zhou)); Mengxuan Zhang (Australian National University); Wen Hua (The Hong Kong Polytechnic University); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- CAGRA: Highly Parallel Graph Construction and Approximate Nearest Neighbor Search for GPUs
  Hiroyuki Ootomo (NVIDIA); Akira Naruse (NVIDIA); Corey Nolet (NVIDIA); Ray Wang (NVIDIA); Tamas Feher (NVIDIA); Yong Wang (NVIDIA)
- VisionEmbedder: Bit-Level-Compact Key-Value Storage with Constant Lookup, Rapid Updates, and Rare Failure
  Yuhan Wu (Peking University); Feiyu Wang (Peking University); Yifan Zhu (Peking university); Zhuochen Fan (Peking University); zhi ting xiong (NUDT); Tong Yang (Peking University); Bin Cui (Peking University)
- Efficient Reverse k Approximate Nearest Neighbor Search over High-Dimensional Vectors
  Yitong Song (Shanghai Jiao Tong University); Kai Wang (Shanghai Jiao Tong University); Bin Yao (Shanghai Jiao Tong University); Zhida Chen (Alibaba Group); Jiong Xie (Alibaba Group); Feifei Li (Alibaba Group)
- HJG: An Effective Hierarchical Joint Graph for ANNS in Multi-Metric Spaces
  Yifan Zhu (Zhejiang University); Lu Chen (Zhejiang University); Yunjun Gao (Zhejiang University); Ruiyao Ma (Zhejiang University); Baihua Zheng (Singapore Management University); Jingwen Zhao (Huawei)

Graphs, Networks, and Semistructured Data XV - [In Theater 10, Chair: Chenxu Wang]

- SACH: Significant-attributed Community Search in Heterogeneous Information Networks
  Yanghao Liu (Institute of Computing Technology, Chinese Academy of Sciences); Fangda Guo (Institute of Computing Technology, Chinese Academy of Sciences); Bingbing Xu (Institute of Computing Technology,University of Chinese Academy of Sciences); Peng Bao (Beijing Jiaotong University); Huawei Shen (Institute of Computing Technology, Chinese Academy of Sciences); Xueqi Cheng (Institute of Computing Technology, Chinese Academy of Sciences)
- TimeSGN: Scalable and Effective Temporal Graph Neural Network
  Yuanyuan Xu (University of New South Wales); Wenjie Zhang (University of New South Wales); Ying Zhang (University of Technology Sydney); Maria Orlowska (Polish-Japanese Institute of Information Technology); Xuemin Lin (Shanghai Jiaotong University)
- Variable-length Path Query Evaluation based on Worst-Case Optimal Joins
  Mingdao Li (Hunan University); Peng Peng (Hunan University); Zheyuan Hu (Hunan University); Lei Zou (Peking University); Zheng Qin (Hunan University)
- NewSP: A New Search Process for Continuous Subgraph Matching over Dynamic Graphs
  li ziming (Hunan University); Youhuan Li (Hunan University); Xinhuan Chen (Tencent Inc.); Lei Zou (Peking University); Yang Li (Tencent Inc.); Xiaofeng Yang (Tencent); Hongbo Jiang (Hunan University)
- Querying Cohesive Subgraph regarding Span-Constrained Triangles on Temporal Graphs
  Chuhan Hu (Wuhan University); Ming Zhong (Wuhan University); Yuanyuan Zhu (Wuhan University); Tieyun Qian (Wuhan University); Ting Yu (Zhejiang Lab); Hongyang Chen (Zhejiang Lab); Mengchi Liu (South China Normal University); Jeffrey Xu Yu (Chinese University of Hong Kong)
- Generating Robust Counterfactual Witnesses for Graph Neural Networks
  Dazhuo Qiu (Aalborg University); Mengying Wang (Case Western Reserve University); Arijit Khan (Aalborg University); Yinghui Wu (Case Western Reserve University)
- Generative and Contrastive Paradigms Are Complementary for Graph Self-Supervised Learning
  Yuxiang Wang (Wuhan University); Xiao Yan (Centre for Perceptual and Interactive Intelligence (CPII)); Chuang Hu (Wuhan University); Quanqing Xu (OceanBase, Ant Group); Chuanhui Yang (OceanBase); Fangcheng Fu (Peking University); Wentao Zhang (Peking University); Hao Wang (Wuhan University); Bo Du (Wuhan University); Jiawei Jiang (Wuhan University)

Data Mining and Knowledge Discovery XVI - [In Theater 9, Chair: TBD]

- LightLT: a Lightweight Representation Quantization Framework for Long-tail Data
  Haoyu Wang (Purdue University); Ruirui Li (Amazon); Zhengyang Wang (Amazon); Xianfeng Tang (Amazon); Danqing Zhang (Amazon); Monica Cheng (Amazon); Bing Yin (Amazon); Jasha Droppo (Amazon); Suhang Wang (Pennsylvania State University); Jing Gao (Purdue University)
- TSec: An Efficient and Effective Framework for Time Series Classification
  Yuanyuan Yao (Zhejiang University); Hailiang Jie (Zhejiang University); Lu Chen (Zhejiang University); Tianyi Li (Aalborg University); Yunjun Gao (Zhejiang University); Shi-ting Wen (NingboTech University)
- McCatch: Scalable Microcluster Detection in Dimensional and Nondimensional Datasets
  Braulio Sánchez (University of São Paulo); Robson Cordeiro (University of São Paulo); Christos Faloutsos (CMU)
- Contrastive Learning for Fraud Detection from Noisy Labels
  Vinay Madanbhavi Shashidhar (University of Arkansas); Shuhan Yuan (Utah State University); Xintao Wu (University of Arkansas)
- Adapting Large Language Models by Integrating Collaborative Semantics for Recommendation
  Bowen Zheng (Renmin University of China); Yupeng Hou (UC San Diego); Hongyu Lu (Tencent); Yu Chen (Tencent); Wayne Xin Zhao (Renmin University of China); Ming Chen (Tencent); Ji-Rong Wen (Renmin University of China)
- Effective Data Selection and Replay for Unsupervised Continual Learning
  Hanmo Liu (Hong Kong University of Science and Technology (Guang Zhou)); Shimin Di (The Hong Kong University of Science and Technology); Haoyang Li (The Hong Kong University of Science and Technology); Shuangyin Li (South China Normal University); Lei Chen (Hong Kong University of Science and Technology); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- Target-agnostic Source-free Domain Adaptation for Regression Tasks
  Tianlang HE (The Hong Kong University of Science and Technology); Zhiqiu Xia (The Hong Kong University of Science and Technology); Jierun Chen (The Hong Kong University of Science and Technology); Haoliang Li (CityU); S.-H. Gary Chan (The Hong Kong University of Science and Technology)

Spatial Databases and Temporal Databases IV - [In Theater 4, Chair: Lei Li]

- FRESH: Towards Efficient Graph Queries in an Outsourced Graph
  Kai Huang (Macau University of Science and Technology); LI Yunqi (HKUST); Qingqing Ye (Hong Kong Polytechnic University); Yao Tian (The Hong Kong University of Science and Technology); Xi Zhao (HKUST); Yue Cui (The Hong Kong University of Science and Technology); Haibo Hu (Hong Kong Polytechnic University); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- Managing the Future: Route Planning Influence Evaluation in Transportation Systems
  Zizhuo Xu (The Hong Kong University of Science and Technology); Lei Li (The Hong Kong University of Science and Technology (Guang Zhou)); Mengxuan Zhang (Australian National University); Yehong Xu (Hongkong university of science and technology); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- Scalable Distance Labeling Maintenance and Construction for Dynamic Small-World Networks
  Xinjie Zhou (The Hong Kong University of Science and Technology); Mengxuan Zhang (Australian National University); Lei Li (The Hong Kong University of Science and Technology (Guang Zhou)); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- Congestion-mitigating Spatiotemporal Routing in Road Networks
  Libin Wang (Hong Kong University of Science and Technology); Raymond Chi-Wing Wong (Hong Kong University of Science and Technology); Christian S. Jensen (Aalborg University)
- A Just-In-Time Framework for Continuous Routing
  Jing ZHAO (HKUST); Lei Li (The Hong Kong University of Science and Technology (Guang Zhou)); Mengxuan Zhang (Australian National University); Zihan Luo (The Hong Kong University of Science and Technology (Guang Zhou)); Xi Zhao (HKUST); Xiaofang Zhou (The Hong Kong University of Science and Technology)
- QSRP: Efficient Reverse k -Ranks Query Processing on High-dimensional Embeddings
  Zheng Bian (The Hong Kong Polytechnic University); Xiao Yan (Centre for Perceptual and Interactive Intelligence (CPII)); JIAHAO ZHANG (The Hong Kong Polytechnic University); Man Lung Yiu (Hong Kong Polytechnic University); Bo Tang (Southern University of Science and Technology)
- FedCTQ: A Federated-based Framework for Accurate and Efficient Contact Tracing Query
  Zhihao Zeng (Zhejiang University); Ziquan Fang (Zhejiang University); Lu Chen (Zhejiang University); Yunjun Gao (Zhejiang University); Kai Zheng (University of Electronic Science and Technology of China); Gang Chen (Zhejiang University)

Information Integration and Data Quality V - [In Theater 3, Chair: Cristian Consonni]

- Dual-Teacher De-biasing Distillation Framework for Multi-domain Fake News Detection
  Jiayang Li (Jinan university); Xuan Feng (Jinan University); Tianlong Gu (Jinan University); Liang Chang (Guilin University of Electronic Technology)
- Are There Fundamental Limitations in Supporting Vector Data Management in Relational Databases? A Case Study of PostgreSQL
  Yunan Zhang (Purdue University - West Lafayette); Shige Liu (Purdue University); Jianguo Wang (Purdue University)
- Compression and In-Situ Query Processing for Fine-Grained Array Lineage
  Jinjin Zhao (University of Chicago); Sanjay Krishnan (U Chicago)
- TSDDISCOVER: Discovery Data Dependency for Time Series Data
  Xiaoou Ding (Harbin Institute of Technology); Yingze Li (HIT); Hongzhi Wang (Harbin Institute of Technology); Chen Wang (" Tsinghua University, China"); Yida Liu (Harbin Institute of Technology); Jianmin Wang ("Tsinghua University, China")
- Time Series Data Cleaning under Expressive Constraints on both Rows and Columns
  Xiaoou Ding (Harbin Institute of Technology); Genglong Li (Harbin Institute of Technology); Hongzhi Wang (Harbin Institute of Technology); Chen Wang (" Tsinghua University, China"); Yichen Song (HIT)
- Cost-Effective In-Context Learning for Entity Resolution: A Design Space Exploration
  Meihao Fan (Renmin University of China); Xiaoyue Han (Renmin University of China); Ju Fan (Renmin University of China); Chengliang Chai (Beijing Institute of Technology); Nan Tang (HKUST (GZ)); Guoliang Li (Tsinghua University); Xiaoyong Du (Renmin University of China)
- A Multi-Task Learning Framework for Reading Comprehension of Scientific Tabular Data
  Xu Yang (Beijing Institute of Technology); Meihui Zhang (Beijing Institute of Technology); Ju Fan (Renmin University of China); Zeyu Luo (Beijing Institute of Technology); Yuxin Yang (Beijing Institute of Technology)

Data Mining and Knowledge Discovery XVIII [In Theater 2, Chair: TBD]

- AdapTraj: A Multi-Source Domain Generalization Framework for Multi-Agent Trajectory Prediction
  Tangwen Qian (Institute of Computing Technology, Chinese Academy of Sciences); Yile Chen (Nanyang Technological University); Gao Cong (Nanyang Technological Univesity); yongjun xu (Institute of Computing Technology, Chinese Academy of Sciences); Fei Wang (Institute of Computing Technology, Chinese Academy of Sciences)
- Towards Effective Next POI Prediction: Spatial and Semantic Augmentation with Remote Sensing Data
  Nan Jiang (Beijing University of Posts and Telecommunications); Haitao Yuan (Nanyang Technological University); Jianing Si (Beijing University of Posts and Telecommunications); Minxiao Chen (Beijing University of Posts and Telecommunications); Shangguang Wang (State Key Laboratory of Networking and Switching Technology)
- KartGPS: Knowledge Base Update with Temporal Graph Pattern-based Semantic Rules
  Hao Xin (Hong Kong University of Science and Technology); Lei Chen (Hong Kong University of Science and Technology)
- Optimizing Probabilistic Box Embeddings with Distance Measures
  Lang Mei (Renmin University of China); Jiaxin Mao (Renmin University of China); Ji-Rong Wen (Renmin University of China)
- A Multi-View Clustering Algorithm for Short Text
  Minkuan Lu (Shandong University); Jianhua Yin (Shandong University); Kaijun Wang (HBIS Digital Technology Co.,ltd); Liqiang Nie (Harbin Institute of Technology (Shenzhen))

A Comprehensive Tutorial on the over 100 years of Diagrammatic Representations of Logical Statements and Relational Queries (Part B) [In Theater 1]

Abstract:
Query formulation is increasingly performed by systems that need to guess a user's intent (e.g. via spoken word interfaces). But how can a user know that the computational agent is returning answers to the "right" query? More generally, given that relational queries can become pretty complicated, how can we help users understand existing relational queries, whether human-generated or automatically generated? Now seems the right moment to revisit a topic that predates the birth of the relational model: developing visual metaphors that help users understand relational queries.
This lecture-style tutorial surveys the key visual metaphors developed for diagrammatic representations of logical statements and relational expressions across both the relational database community and the much older diagrammatic reasoning community. We will survey the history and state-of-the art of relationally-complete diagrammatic representations of relational queries, discuss the key visual metaphors developed in over a century of investigating diagrammatic languages, and organize the landscape by mapping their used visual alphabets to the syntax and semantics of Relational Algebra (RA) and Relational Calculus (RC). Tutorial website: https://northeastern-datalab.github.io/diagrammatic-representation-tutorial/

Presenters:

Wolfgang Gatterbauer is an Associate Professor at the Khoury College of Computer Sciences at Northeastern University. His research interests lie in the intersection of theory and practice of data management. He received an NSF Career award and -- with his students and collaborators -- a best paper award at EDBT 2021, best-of-conference mentions for PODS 2021, SIGMOD 2017, WALCOM 2017, and VLDB 2015, and two out of three reproducibility awards for papers published at SIGMOD 2020.

Poster Session - [In Speys]

Posters of the 14:00 Session

Poster Session - [In B2B]

Posters of the 14:00 Session

Notes

Notes

Notes

Notes

Notes