

43rd International Conference on Very Large Data Bases



28 August – 1 September 2017 · Munich, Germany



www.vldb.org/2017

Welcome Message

VLDB is a premier annual international forum for data management and database researchers, vendors, practitioners, application developers, and users. The annual conference consists of a mix of research talks, tutorials, demonstrations, and workshops. Its topical coverage includes current issues in data management, database, and information systems research. Data management and databases remain among the main technological cornerstones of the applications of the twenty-first century. With the emergence of Big Data, data-related technologies are becoming more important than ever before.

VLDB 2017 is taking place at the Technical University of Munich (TUM), one of Europe's top universities. It is committed to excellence in research and teaching, interdisciplinary education and the active promotion of promising young scientists. The university also forges strong links with companies and scientific institutions across the world. TUM was one of the first universities in Germany to be named a University of Excellence. Moreover, TUM regularly ranks among the best European universities in international rankings. The VLDB conference takes place at TUM's (original) downtown campus whereas most natural science and technical faculties, such as mechanical engineering, computer science, mathematics, chemistry, and physics are lo-

cated at the modern Garching campus, about 15 km north of Munich. The VLDB 2017 conference is held as part of TUM’s 50 year anniversary of establishing Computer Science as a scientific discipline in Munich. Since TUM was founded in 1868 (almost 150 years ago) it contributed significantly to the conversion of Bavaria from an agricultural to a highly industrialized region and also to the overall technological progress.

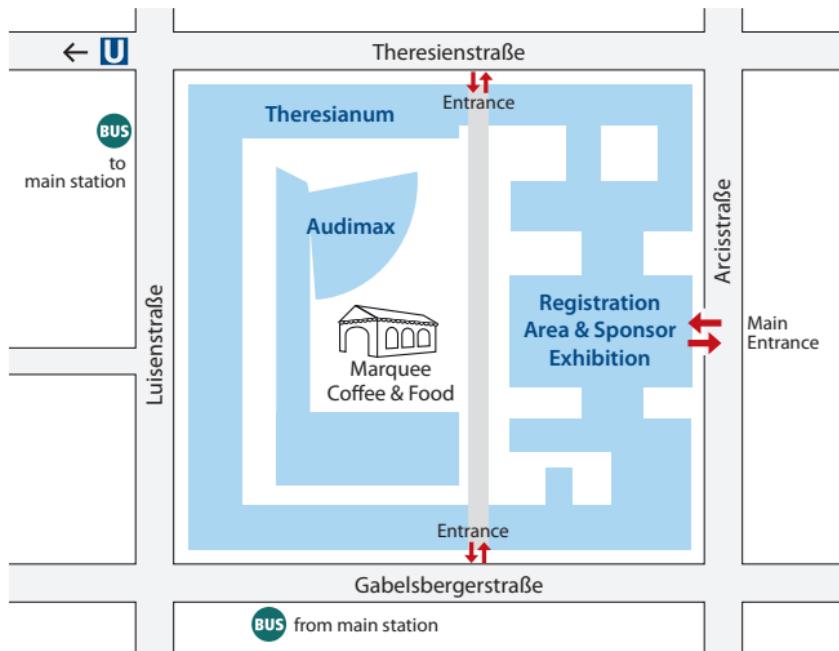
VLDB 2017 received almost 750 research paper submissions, of which 133 have been accepted for presentation at the conference. The conference program also includes 8 “roll-over” papers from VLDB 2016, for a total of 141 research papers. The full conference program also includes 20 papers from the industrial track, 35 research demonstrations, 8 tutorials, and a panel on cross-disciplinary research. Each day of the conference also features a plenary keynote presentation. The main conference is flanked by a variety of workshops on topics of particular interest to the community.

The VLDB 2017 technical program reflects the efforts of hundreds of members of the research community who have prepared papers, demonstrations and presentations for the conference. It is also the result of more than a year of work by the conference officers and the PVLDB Review Board. We’re looking forward to an outstanding program, and we hope that you enjoy it!

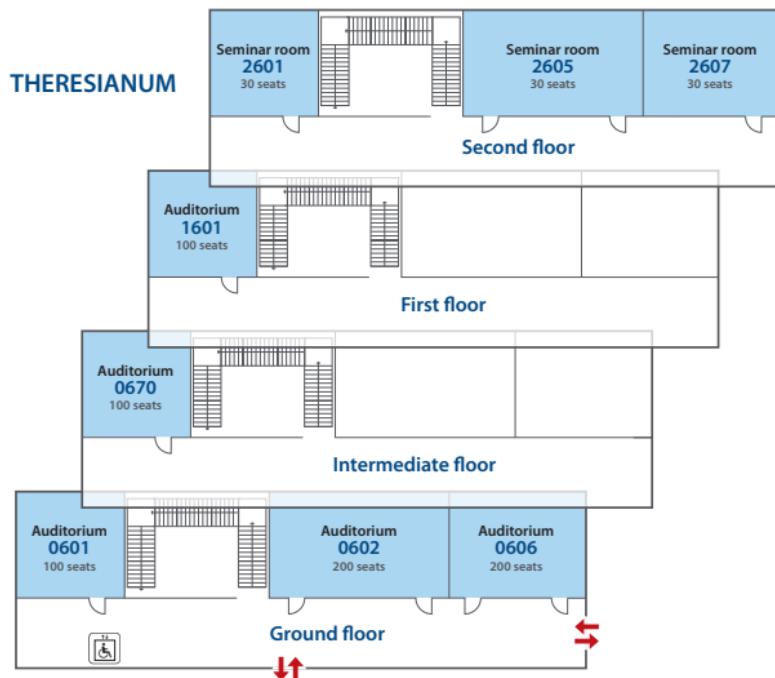
Alfons Kemper, TUM
Thomas Neumann, TUM
VLDB 2017 General Chairs

Peter Boncz, CWI
Ken Salem, University of Waterloo
VLDB 2017 Program Committee Chairs

Location TUM



Lecture Halls in the Theresianum



Street Map

Conference Location (A)

TUM, Arcisstraße 21

Reception (1)

Old Town Hall, Marienplatz 15

Banquet (B)

Hofbräuhaus, Platzl 9



Map data (c) OpenStreetMap contributors

Social Events

Both social events are close by the central place in Munich (Marienplatz). You can either walk there as shown on the map (2.5 km, about 30 min), or take public transport.

Public transport options:

- bus 100 (leaves from Gabelsbergerstraße, south entrance of TUM, direction Ostbahnhof) to Odeonsplatz. There, switch to the subway U3 or U6 towards Marienplatz (1 stop).
- live routing: <https://goo.gl/1T1Sg6>

The **Old Town Hall** is next to the Marienplatz, 50m to the east. It is a white building (<https://goo.gl/XE2OUV>), not to be confused with the large, red New Town Hall immediately next to the Marienplatz.

The **Hofbräuhaus** (<https://goo.gl/SUUQvW>) is a few minutes east and north of the Marienplatz, as shown on the map. Turn left behind the Old Town Hall into Sparkassenstraße, right into Münzstraße, and left into Platzl.

WLAN

Preferably, just use the *EDUROAM* network if you have eduroam access. Otherwise, follow the instructions below:



Wi-Fi-Guide for mwn-events

Wi-Fi name (SSID): mwn-events

Username: VLDB2017

Password: YTn06kdF

Valid from Fri Aug 25 06:00 2017 to Sat Sep 9
23:59 2017

Configuration profiles for wireless network access are available via the QR code or this URL:

<https://www.lrz.de/wlan> (follow the link mwn-events) Access to this site is available via the open Wi-Fi (the SSID) "lrz".

Our Sponsors

Platinum Sponsors



Gold Sponsors



Silver Sponsors



Bronze Sponsors



Exhibitors



12 Program at a Glance

Tuesday, August 29							
Time	Audimax	602	606	601	670	1601	260*
08:30-10:00	Welcome Messages + Wolfgang Lehner keynote						
10:00-10:30	Coffee Break (tent)						
10:30-12:00	Panel	DB Engines 1	Data Cleaning	Spatial Data Management 1	Graphs and Networks 1	Tutorial 1	Demo Group A
12:00-13:30	Lunch (tent)						
13:30-15:00	Query Processing and Optimization	Stream Processing 1	Data Formats	Privacy and Security	Event Processing	Tutorial 2	Demo Group B
15:00-15:30	Coffee Break (tent)						
15:30-17:00	Transactions	Spatial Data management 2	Graphs and Networks 2	Information Integration	Applications	Tutorial 3	Demo Group C
17:00-18:00	Poster Reception (Foyer Audimax)						

Thursday, August 31							
Time	Audimax	602	606	601	670	1601	260*
08:30-10:30	Endowment Awards + Demo Award + Jens Dittrich plenary						
10:30-10:55	Coffee Break (tent)						
10:55-12:00	Transactions and Persistence	Data Access	Data Statistics	Potpourri			
12:00-13:30	Lunch (tent)						
13:30-15:00	Data Partitioning	Graphs and Networks 5	Visualizat ion	Distributed Systems and Cloud 1	Estimation and Approximation	Tutorial 7	Demo Group D
15:00-15:30	Coffee Break (tent)						
15:30-17:00	DB Engines 2	Query Processing	Text and Semi- Structured	Spatial Data Management 3	Distributed Systems and Cloud 2	Tutorial 8	Demo Group B
17:00-18:00	Poster Reception (Foyer Audimax)						

Friday, September 1

Time	Audimax	602	606	601	670	1601	260*
08:30-10:00		MATES	ADMS	DMAH	DBPL	BOSS	BOSS
10:00-10:30				Coffee Break (tent)			
10:30-12:00		MATES	ADMS	DMAH	DBPL	BOSS	BOSS
12:00-13:30				Lunch (tent)			
13:30-15:00		MATES	ADMS	DMAH	DBPL	BOSS	BOSS
15:00-15:30				Coffee break (tent)			
15:30-17:00		MATES	ADMS	DMAH	DBPL	BOSS	BOSS

Tuesday 08/29/2017 08:30-10:00

Welcome Messages + Wolfgang Lehner
keynote

Location: Audimax

Overview General Chairs (Kemper, Neumann), Welcome Hans-Joachim Bungartz (Dean of Informatics, TUM), Program Overview PC Chairs (Boncz, Salem)

The Data Center under your Desk - How Disruptive is Modern Hardware for DB System Design?

Wolfgang Lehner (Technische Universität Dresden)

While we are already used to see more than 1,000 cores within a single machine, the next processing platforms for database engines will be widely heterogeneous with built-in GPU-style processors as well as specialized FPGAs and chips with domain-specific instruction sets taking advantage of the “Dark Silicon” effect. Moreover, the traditional volatile as well as the upcoming non-volatile RAM with capacities in the 100s of TBytes per machine will provide great opportunities for storage engines but also call for radical changes on the architecture of such systems. Finally, the emergence of economically affordable, high-speed/low-latency interconnects as a basis for rack-scale computing is questioning long-standing folklore algorithmic assumptions but will certainly play

an important role of the big picture for building modern data management platforms. While database research on modern hardware has already produced a rich bouquet of promising results targeting a wide variety of hardware directions, the talk will try to classify and review existing approaches from a performance, robustness, as well as energy efficiency perspective. Moreover, the talk will discuss the overall question on how these results can be incorporated into the design and implementation of modern DB systems. The goal is therefore to outline current trends and research activities as well as to pinpoint to interesting starting points for further research activities.

Wolfgang Lehner is full professor and head of the Database Technology Group as well as director of the Institute for System Architecture at TU Dresden, Germany. His research focuses on database system architecture specifically looking at crosscutting aspects from algorithms down to hardware-related aspects in main-memory centric settings. He is part of TU Dresden's research cluster of excellence with topics in energy-aware computing, resilient data structures on unreliable hardware, and orchestration of widely heterogeneous systems. He is heading a Research Training Group on large-scale adaptive system software design and acts as a principal investigator in Germany's national "Compe-

tence Center for Scalable Data Services and Solutions” (ScaDS). Wolfgang also maintains a close research relationship with the SAP HANA development team. He serves the community in many PCs, is an elected member of the VLDB Endowment, is chairing the review board of Computer Science within the German Research Foundation (DFG), and is an appointed member of the Academy of Europe.

Tuesday 08/29/2017 10:00-10:30

Coffee Break

Location: Tent

Tuesday 08/29/2017 10:30-12:00

Panel

Location: Audimax

Interdisciplinary research and the impact of data management/systems research outside our own community

Timos Sellis (moderator) (Swinburne University of Technology), Michael Franklin (University of Chicago), Johann-Christoph Freytag (Humboldt-Universität zu Berlin), Raymond Ng (University of British Columbia), Matthias Renz (George Mason University), Kian-Lee Tan (NUS)

Location: Auditorium 602

Chair: Ippokratis Pandis

Resumable Online Index Rebuild in SQL Server (industrial)

Panagiotis Antonopoulos (Microsoft), Hanuma Kodavalla (Microsoft), Alex Tran (Microsoft), Nitish Upreti (Microsoft), Chaitali Shah (Microsoft), Mirek Sztajno (Microsoft)

Quaestor: Query Web Caching for Database-as-a-Service Providers (industrial)

Felix Gessert (Baqend GmbH), Michael Schaarschmidt (University of Cambridge), Wolfram Wingerath (Universität Hamburg), Erik Wiit (Baqend GmbH), Eiko Yoneki (University of Cambridge), Norbert Ritter (University of Hamburg)

BlueCache: A Scalable Distributed Flash-based Key-value Store

Shuotao Xu (MIT), Sungjin Lee (Inha University), Sang-Woo Jun (Massachusetts Institute of Technology), Ming Liu (Massachusetts Institute of Technology), Jamey Hicks (Accelerated Tech), Arvind (MIT)

Quill: Efficient, Transferable, and Rich Analytics at Scale

Badrish Chandramouli (Microsoft Research), Raul Castro Fernandez (MIT), Jonathan Goldstein (Microsoft Research), Ahmed Eldawy (University of Minnesota), Abdul Quamar (University of Maryland)

Data Cleaning

Location: Auditorium 606

Chair: Mourad Ouzzani

HoloClean: Holistic Data Repairs with Probabilistic Inference

Theodoros Rekatsinas (Stanford University), Xu Chu (University of Waterloo), Ihab Ilyas (University of Waterloo), Chris Re (Stanford University)

CleanM: An Optimizable Query Language for Unified Scale-Out Data Cleaning

Stella Giannakopoulou (EPFL), Manos Karpathiotakis (EPFL), Benjamin Gaidioz (EPFL), Anastasia Ailamaki (EPFL)

ZooBP: Belief Propagation for Heterogeneous Networks

Dhivya Eswaran (CMU), Stephan Guennemann (TUM), Christos Faloutsos (CMU), Disha Makhija (Flipkart), Mohit Kumar (Flipkart)

Time Series Data Cleaning: From Anomaly Detection to Anomaly Repairing

Aoqian Zhang (Tsinghua University), Shaoxu Song (Tsinghua University), Jianmin Wang (Tsinghua University), Philip Yu (University of Illinois at Chicago)

Spatial Data Management 1

Location: Auditorium 601

Chair: Reynold Cheng

Flexible Online Task Assignment in Real-Time Spatial Data

Yongxin Tong (Beihang University), Libin Wang (Beihang University), Zimu Zhou (ETH), Bolin Ding (Microsoft Research), Lei Chen (HKUST), Jieping Ye (Didi Research), Ke Xu (Beihang University)

Path Cost Distribution Estimation Using Trajectory Data

Jian Dai (NUS), Bin Yang (AAU), Chenjuan Guo (AAU), Christian Jensen (Aalborg University), Jilin Hu (AAU)

VIP-Tree: An Effective Index for Indoor Spatial Queries

Zhou Shao (Monash University), Muhammad Cheema (Monash University), David Taniar (Monash University), Hua Lu (Aalborg University)

One-Pass Error Bounded Trajectory Simplification

Xuelian Lin (Beihang University), Shuai Ma (Beihang

University), Han Zhang (Beihang University), Tianyu Wo (Beihang University), Jinpeng Huai (Beihang University)

Graphs and Networks 1

Location: Auditorium 670

Chair: Semih Salihoglu

Scalable Distributed Subgraph Enumeration

Longbin Lai (CSE), Lu Qin (QCIS), Xuemin Lin (CSE), Ying Zhang (QCIS), Lijun Chang (CSE)

Fast Algorithm for the Lasso based L1-Graph Construction

Yasuhiro Fujiwara (NTT), Yasutoshi Ida (NTT), Junya Arai (NTT), Mai Nishimura (NTT), Sotetsu Iwamura (NTT)

When Engagement Meets Similarity: Efficient (k,r)-Core Computation on Social Networks

Fan Zhang (UTS), Ying Zhang (QCIS), Lu Qin (QCIS), Wenjie Zhang (CSE), Xuemin Lin (CSE)

Tutorial 1

Location: Auditorium 1601

Caching at the Web Scale

Victor Zakhary (UCSB), Amr El Abbadi (UCSB), Divyakant Agarwal (UCSB)

Location: Classrooms 2601, 2605, 2607

A Confidence-Aware Top-k Query Processing Toolkit on Crowdsourcing

Yan Li (University of Macau), Ngai Meng Kou (University of Macau), Hao Wang (Nanjing University), Leong Hou U (University of Macau), Zhiguo Gong (University of Macau)

DataTweener: A Demonstration of a Tweening Engine for Incremental Visualization of Data Transforms

Meraj Ahmed Khan (Ohio State University), Larry Xu (UC Berkeley), Arnab Nandi (Ohio State University), Joseph Hellerstein (UC Berkeley)

A Demonstration of Stella: A Crowdsourcing-Based Geotagging Framework

Christopher Jonathan (University of Minnesota), Mohamed Mokbel (University of Minnesota)

Interactive Navigation of Open Data Linkages

Erkang Zhu (University of Toronto), Ken Pu (UOIT), Fateh-meh Nargesian (University of Toronto), Renee Miller (University of Toronto)

noWorkflow: a Tool for Collecting, Analyzing, and Managing Provenance from Python Scripts

*João Felipe Pimentel (Universidade Federal Fluminense),
Leonardo Murta (Universidade Federal Fluminense),
Vanessa Braganholo (Universidade Federal Fluminense),
Juliana Freire (NYU)*

ARShop: A Cloud-based Augmented Reality System for Shopping

Chao Wang (NUS), Yihao Feng (Dartmouth College), Qi Guo (NUS), Zhaoxian Li (NUS), Kexin Liu (NUS), Zijian Tang (NUS), Anthony Tung (NUS), Lifu Wu (NUS), Yuxin Zheng (NUS)

Debugging Transactions and Tracking their Provenance with Reenactment

Xing Niu (IIT), Bahareh Sadat Arab (Illinois Institute of Technology), Seokki Lee (Illinois Institute of Technology), Su Feng (Illinois Institute of Technology), Xun Zou (Illinois Institute of Technology), Dieter Gawlick (Oracle), Vasudha Krishnaswamy (Oracle), Zhen Hua Liu (Oracle), Boris Glavic (Illinois Institute of Technology)

FlashView: An Interactive Visual Explorer for Raw Data

Zhifei Pang (Zhejiang University), Sai Wu (Zhejiang University), Gang Chen (Zhejiang University), Ke Chen (Zhejiang University), Lidan Shou (Zhejiang University)

Automating Data Citation in CiteDB

Abdussalam Alawini (University of Pennsylvania), Susan

Davidson (University of Pennsylvania), Wei Hu (University of Pennsylvania), Yinjun Wu (University of Pennsylvania)

Tuesday 08/29/2017 12:00-13:30

Lunch

Location: Tent

Tuesday 08/29/2017 13:30-15:00

Query Processing and Optimization

Location: Audimax

Chair: Fatma Ozcan

Distributed Join Algorithms on Thousands of Cores

Claude Barthels (ETH), Gustavo Alonso (ETH), Torsten Hoefler (ETH), Timo Schneider (ETH), Ingo Müller (ETH)

Adaptive Work Placement for Query Processing on Heterogeneous Computing Resources

Tomas Karnagel (TU Dresden), Dirk Habich (TU Dresden), Wolfgang Lehner (TU Dresden)

Automatic Algorithm Transformation for Efficient Multi-Snapshot Analytics on Temporal Graphs

Manuel Then (TUM), Timo Kersten (TUM), Stephan Guenemann (TUM), Alfons Kemper (TUM), Thomas Neumann

Towards Linear Algebra over Normalized Data

Lingjiao Chen (UW-Madison), Arun Kumar (University of California), Jeffrey Naughton (Google), Jignesh Patel (UW-Madison)

Stream Processing 1

Location: Auditorium 602

Chair: Kai-Uwe Sattler

State Management in Apache Flink®: Consistent Stateful Distributed Stream Processing (industrial)

Paris Carbone (KTH), Stephan Ewen (Data Artisans), Gyula Fóra (King Digital Entertainment Limited), Seif Haridi (KTH), Stefan Richter (Data Artisans), Kostas Tzoumas (Data Artisans)

Dhalion:Self-Regulating Stream Processing in Heron (industrial)

Avrilia Floratou (Microsoft), Ashvin Agrawal (Microsoft), Bill Graham (Twitter), Sriram Rao (Microsoft), Karthik Ramasamy (Twitter)

A Declarative Query Processing System for Now-casting

Dolan Antenucci (University of Michigan), Michael Anderson (University of Michigan), Michael Cafarella (University of Michigan)

Samza: Stateful Scalable Stream Processing at LinkedIn (industrial)

Shadi A Noghabi (University of Illinois at Urbana-Champaign), Kartik Paramasivam (LinkedIn), Yi Pan (LinkedIn), Navina Ramesh (LinkedIn), Jon Bringhurst (LinkedIn), Indranil Gupta (UIUC), Roy Campbell (University of Illinois at Urbana-Champaign)

Data Formats

Location: Auditorium 606

Chair: Holger Pirk

Mison: A Fast JSON Parser for Data Analytics

Yinan Li (Microsoft Research), Nikos R. Katsipoulakis (University of Pittsburgh), Badrish Chandramouli (Microsoft Research), Jonathan Goldstein (Microsoft Research), Donald Kossmann (ETH)

FAD.js: Fast JSON Data Access Using JIT-based Speculative Optimizations (industrial)

Daniele Bonetta (Oracle Labs), Matthias Brantner (Oracle Labs)

MILC: Inverted List Compression in Memory

Jianguo Wang (UCSD), Chunbin Lin (UCSD), Ruining He (UCSD), Moojin Chae (UCSD), Yannis Papakonstantinou (UCSD), Steven Swanson (UCSD)

Don't Hold My Data Hostage - A Case For Client

Protocol Redesign

Mark Raasveldt (CWI), Hannes Mühleisen (CWI)

Privacy and Security

Location: Auditorium 601

Chair: sharad mehrotra

Privacy-preserving Network Provenance

Yuankai Zhang (Georgetown University), Adam O'Neill (Georgetown University), Micah Sherr (Georgetown University), Wenchao Zhou (Georgetown University)

Plausible Deniability for Privacy-Preserving Data Synthesis

Vincent Bindschaedler (UIUC), Reza Shokri (CornellTech), Carl Gunter (UIUC)

Understanding the Sparse Vector Technique for Differential Privacy

Min Lyu (University of Science and Technology of China), Dong Su (Purdue University), Ninghui Li (Purdue University)

SMCQL: Secure Query Processing for Private Data Networks

Johes Bater (Northwestern University), Greg Elliott (Northwestern University), Craig Eggen (Northwestern University), Satyender Goel (Northwestern University), Abel Kho (Northwestern University), Jennie Rogers

(Northwestern University)

Event Processing

Location: Auditorium 670

Chair: Bernhard Seeger

IL-Miner: Instance-Level Discovery of Complex Event Patterns

Lars George (Humboldt-Universität zu Berlin), Bruno Cadonna (Humboldt-Universität zu Berlin), Matthias Weidlich (Humboldt-Universität zu Berlin)

Truth Discovery for SpatioTemporal Events from Crowdsourced Data

Daniel Garcia Ulloa (Emory University), Li Xiong (Emory University), Vaidy Sunderam (Emory University)

Computing Longest Increasing Subsequences over Sequential Data Streams

Youhuan Li (Peking University), Lei Zou (Peking University), Huaming Zhang (University of Alabama in Huntsville), Dongyan Zhao (Peking University)

Stochastic Data Acquisition for Answering Queries as Time Goes by

Zheng Li (University of Massachusetts Lowell), Tingjian Ge (University of Massachusetts Lowell)

Tutorial 2

Location: Auditorium 1601

Summarizing Static and Dynamic Big Graphs

Arijit Khan (NTU Singapore), Sourav S Bhowmick (Nanyang Technological University), Francesco Bonchi (ISI Foundation)

Demo Group B

Location: Classrooms 2601, 2605, 2607

Thoth in Action: Memory Management in Modern Data Analytics

Mayuresh Kunjir (Duke University), Shivnath Babu (Duke University)

Monopedia: Staying Single is Good Enough - The HyPer Way for Web Scale Applications

Maximilian Schüle (TUM), Pascal Schliski (TUM), Thomas Hutzelmann (TUM), Tobias Rosenberger (TUM), Viktor Leis (TUM), Dimitri Vorona (TUM), Alfons Kemper (TUM), Thomas Neumann (TUM)

Dima: A Distributed In-Memory Similarity-Based Query Processing System

Ji Sun (Tsinghua University), Zeyuan Shang (Tsinghua University), Guoliang Li (Tsinghua University), Dong Deng (MIT), Zhifeng Bao (RMIT University)

A BAD Demonstration: Towards Big Active Data

Steven Jacobs (University of California at Riverside), Md Yusuf Sarwar Uddin (University of California at Irvine), Michael Carey (University of California at Irvine), Vagelis Hristidis (University of California at Riverside), Vassilis Tsotras (University of California at Riverside), Nalini Venkatasubram (University of California at Irvine), Yao Wu (University of California at Irvine), Syed Safir (University of California at Irvine), Purvi Kaul (University of California at Irvine), Xikui Wang (University of California at Irvine), Mohiuddin Abdul Qader (University of California), Yawei Li (University of California at Riverside)

A Demonstration of ST-Hadoop: A MapReduce Framework for Big Spatio-temporal Data

Louai Alarabi (University of Minnesota), Mohamed Mokbel (University of Minnesota)

Creation and Interaction with Large-scale Domain-Specific Knowledge Bases

Shreyas Bharadwaj (IBM Watson), Laura Chiticariu (IBM Research-Almaden), Marina Danilevsky (IBM Research-Almaden), Samarth Dhingra (IBM Watson), Samved Divekar (IBM Watson), Arnaldo Carreno-Fuentes (IBM Watson), Himanshu Gupta (IBM Research-India), Nitin Gupta (IBM Research-India), Sang-Don Han (IBM Watson), Mauricio Hernandez (IBM Research-Almaden), Howard Ho (IBM Watson), Parag Jain (IBM

Research-India), Salil Joshi (IBM Research-India), Hima Karanam (IBM Research-India), Saravanan Krishnan (IBM Research-India), Rajasekar Krishnamurthy (IBM Research-Almaden), Yunyao Li (IBM Research-Almaden), Satishkumar Manivannan (IBM Watson), Ashish Mittal (IBM Research-India), Fatma Ozcan (IBM Research-Almaden), Abdul Quamar (IBM Research-Almaden), Poornima Raman (IBM Watson), Diptikalyan Saha (IBM Research-India), Karthik Sankaranarayanan (IBM Research-India), Jaydeep Sen (IBM Research-India), Prithviraj Sen (IBM Research-Almaden), Shivakumar Vaithyanathan (IBM Watson), Mitesh Vasa (IBM Watson), Hao Wang (IBM Watson), Huaiyu Zhu (IBM Research-Almaden)

Exploring big volume sensor data with Vroom

Oscar Moll (MIT), Aaron Zalewski (MIT), Sudeep Pillai (MIT), Samuel Madden (MIT), Michael Stonebraker (MIT), Vijay Gadepally (MIT Lincoln Labs)

DITIR: Distributed Index for High Throughput Trajectory Insertion and Real-time Temporal Range Query

Ruichu Cai (Guangdong University of Technology), Zijie Lu (Guangdong University of Technology), Li Wang (Advanced Digital Sciences Center), Zhenjie Zhang (Advanced Digital Sciences Center), Tom Fu (Advanced Digital Sciences Center), Marianne Winslett (University of

Illinois at Urbana-Champaign)

Tuesday 08/29/2017 15:00-15:30

Coffee Break

Location: Tent

Tuesday 08/29/2017 15:30-17:00

Transactions

Location: Audimax

Chair: Nikolaus Augsten

An Evaluation of Distributed Concurrency Control

Rachael Harding (MIT), Dana Van Aken (CMU), Andrew Pavlo (CMU), Michael Stonebraker (MIT)

The End of a Myth: Distributed Transaction Can Scale

Erfan Zamanian (Brown University), Carsten Binnig (Brown University), Tim Kraska (Brown University), Tim Harris (Oracle Labs)

OrpheusDB: Bolt-on Versioning for Relational Databases

Silu Huang (UIUC), Liqi Xu (UIUC), Jialin Liu (Peking University), Aaron J. Elmore (University of Chicago), Aditya Parameswaran (UIUC)

Spatial Data management 2

Location: Auditorium 602

Chair: Bongki Moon

Effective Indexing for Approximate Constrained Shortest Path Queries on Large Road Networks

Sibo Wang (Nanyang Technological University), Xiaokui Xiao (Nanyang Technological University), Yin Yang (Hamad Bin Khalifa University), Wenqing Lin (Qatar Computing Research Institute)

Trajectory Similarity Join in Spatial Networks

Shuo Shang (KAUST), Lisi Chen (Hong Kong Baptist University), Zhewei Wei (RUC), Christian Jensen (Aalborg University), Kai Zheng (Soochow University), Panos Kalnis (KAUST)

Clue-based Spatio-textual Query

Junling Liu (Northeastern University China), Ke Deng (RMIT University), Huanliang Sun (Shenyang Jianzhu University), Yu Ge (Northeastern University China), Xiaofang Zhou (University of Queensland), Christian Jensen (Aalborg University)

Distributed Trajectory Similarity Search

Dong Xie (University of Utah), Feifei Li (University of Utah), Jeff Phillips (University of Utah)

Graphs and Networks 2

Location: Auditorium 606

Chair: Michael Grossniklaus

Attribute-Driven Community Search

Xin Huang (Hong Kong Baptist University), Laks Lakshmanan (UBC)

Finding the maximum clique in massive graphs

Can Lu (CUHK), Jeffrey Yu (CUHK), Hao Wei (CUHK), Yikai Zhang (CUHK)

Multi-Query Optimization for Subgraph Isomorphism Search

Xuguang Ren (Griffith University), Junhu Wang (Griffith University)

Information Integration

Location: Auditorium 601

Chair: Theodoros Rekatsinas

In Search of an Entity Resolution OASIS: Optimal Asymptotic Sequential Importance Sampling

Neil Marchant (University of Melbourne), Benjamin Rubinstein (University of Melbourne)

Knowledge Verification for LongTail Verticals

Furong Li (NUS), Xin Luna Dong (Amazon), Anno Langen (Google), Yang Li (Google)

LDA*: A Robust and Large-scale Topic Modeling System

Lele Yu (Peking University), Bin Cui (Peking University), Ce Zhang (ETH), Yingxia Shao (PKU)

Stitching Web Tables for Improving Matching Quality

Oliver Lehmberg (University of Mannheim), Christian Bizer (University of Mannheim)

Applications

Location: Auditorium 670

Chair: Qiong Luo

Colt: Concept Lineage Tool for Data Flow Metadata Capture and Analysis (industrial)

Kareem Aggour (GE Global Research), Jenny Weisenberg Williams (GE Global Research), Justin McHugh (GE Global Research), Vijay Kumar (GE Global Research)

Matrix Profile IV: Using Weakly Labeled Time Series to Predict Outcomes (industrial)

Chin-Chia Michael Yeh (UC Riverside), Nickolas Kavantzas (Oracle), Eamonn Keogh (UC Riverside)

Developing a Low Dimensional Patient Class Profile in Accordance to Their Respiration-Induced Tumor Motion (industrial)

Rittika Shamsuddin (University of Texas at Dallas), Bal-

*akrishnan Prabhakaran (University of Texas at Dallas),
Amit Sawant (University of Maryland)*

Probabilistic Demand Forecasting at Scale (industrial)

Joos-Hendrik Boese (Amazon), Valentin Flunkert (Amazon), Jan Gasthaus (Amazon), Tim Januschowski (Amazon), Dustin Lange (Amazon), David Salinas (Amazon), Sebastian Schelter (Amazon), Matthias Seeger (Amazon), Bernie Wang (Amazon)

Tutorial 3

Location: Auditorium 1601

Blockchains and Databases

C. Mohan (IBM Almaden Research Center)

Demo Group C

Location: Classrooms 2601, 2605, 2607

C-Explorer: Browsing Communities in Large Graphs

Yixiang Fang (Hong Kong University), Reynold Cheng (Hong Kong University), Siqiang Luo (Hong Kong University), Jiafeng Hu (Hong Kong University), Kai Huang (Hong Kong University)

GRAPE: Parallelizing Sequential Graph Computations

Wenfei Fan (University of Edinburgh and Beihang University), Jingbo Xu (University of Edinburgh and Beihang University), Yinghui Wu (Washington State University), Wenyuan Yu (Beihang University), Jiaxin Jiang (Hong Kong Baptist University)

STEED: An Analytical Database System for TrEE-structured Data

Zhiyi Wang (Chinese Academy of Sciences), Dongyan Zhou (Chinese Academy of Sciences), Shimin Chen (Chinese Academy of Sciences)

Strider: An Adaptive, Inference-enabled Distributed RDF Stream Processing Engine

Xiangnan Ren (ATOS), Olivier Curé (UPEM LIGM - UMR CNRS 8049), Li Ke (ATOS), Jérémie Lhez (UPEM LIGM - UMR CNRS 8049), Badre Belabbess (ATOS), Tendry Randriamalala (ATOS), Yufan Zheng (ATOS), Gabriel Kepeklian (ATOS)

Explaining and Querying Knowledge Graphs by Relatedness

Valeria Fionda (University of Calabria), Giuseppe Pirrò (ICAR-CNR)

TeCoRe: Temporal Conflict Resolution in Knowledge Graphs

Melisachew Chekol (University of Mannheim), Giuseppe Pirrò (ICAR-CNR), Joerg Schoenfisch (University of

Mannheim), Heiner Stuckenschmidt (University of Mannheim)

PICASSO: Exploratory Search of Connected Subgraph Substructures in Graph Databases

Kai Huang (Fudan University), Sourav S Bhowmick (Nanyang Technological University), Shuigeng Zhou (Fudan University), Byron Choi (Hong Kong Baptist University)

Upsortable: Programming TopK Queries Over Data Streams

Julien Subercaze (Lab Hubert Curien), Christophe Gravier (Lab Hubert Curien), Syed Gillani (Lab Hubert Curien), Abderrahmen Kammoun (Lab Hubert Curien), Frédérique Laforest (Lab Hubert Curien)

Tuesday 08/29/2017 17:00-18:00

Poster Reception

Location: Audimax Foyer

Poster Reception

Poster presentations for all papers that were presented that same day, accompanied by some drinks. Additionally, two VLDB Journal papers will present a poster.

Disjoint Interval Partitioning

Francesco Cafagna (University of Zurich), Michael Boehlen (University of Zurich)

AutoG: A Visual Query Autocompletion Framework for Graph Databases

Peipei Yi (Hong Kong Baptist University), Byron Choi (Hong Kong Baptist University), Sourav S. Bhowmick (Hong Kong Baptist University), Jianliang Xu (Hong Kong Baptist University)

Wednesday 08/30/2017 08:30-10:00

Endowment Update + VLDB Awards +
Michael Franklin keynote

Location: Audimax

Big Data Software: What's Next?

Michael Franklin (University of Chicago)

The Big Data revolution has been enabled in part by a wealth of innovation in software platforms for data storage, analytics, and machine learning. The design of Big Data platforms such as Hadoop and Spark focused on scalability, fault-tolerance and performance. As these and other systems increasingly become part of the mainstream, the next set of challenges are becoming clearer. Requirements for performance are changing as workloads evolve to include techniques such as hardware-accelerated deep learning. But more fundamentally, other issues are moving to the forefront. These include ease of use for a wide range of users,

security, concerns about privacy and potential bias in results, and the perennial problems of data quality and integration from heterogeneous sources. Fortunately, the database community has much to say about all of these topics, and can and should take a leading role in addressing them. In this talk, I will give an overview of how we got here, with an emphasis on the development of the Apache Spark system. I will then focus on these emerging issues with an eye towards where the database community can most effectively engage.

MICHAEL J. FRANKLIN is the Liew Family Chair of Computer Science and Sr. Advisor to the Provost for Computation and Data at the University of Chicago where his research focuses on database systems, data analytics, human-in-the-loop computing, and distributed computing systems. Previously he was the Thomas M. Siebel Professor and Chair of Computer Science at UC Berkeley. He co-founded and directed the Algorithms, Machines and People Laboratory (AMPLab), which created industry-changing open source Big Data software such as Apache Spark and BDAS, the Berkeley Data Analytics Stack. Franklin has nearly three decades(!) of experience with Database Systems projects including the Bubba massively parallel DBMS, the SHORE object-oriented DBMS, the TelegraphCQ and Truviso stream processing systems, the TinyDB and HiFi sensor

query processing systems, and Spark/BDAS. He currently serves as a Board Member of the Computing Research Association and on the NSF CISE Advisory Committee. He is an ACM Fellow, a two-time recipient of the ACM SIGMOD “Test of Time” award and received the Outstanding Advisor award from Berkeley’s Computer Science Graduate Student Association.

Wednesday 08/30/2017 10:00-10:30

Coffee Break

Location: Tent

Wednesday 08/30/2017 10:30-12:00

High Performance Query Processing

Location: Audimax

Chair: Boris Glavic

Bridging the Gap between HPC and Big Data frameworks

Michael Anderson (Intel Labs), Shaden Smith (University of Minnesota), Narayanan Sundaram (Intel), Mihai Capotă (Intel Labs), Zheguang Zhao (Brown University), Subramanya Dulloor (Intel Labs), Nadathur Satish (Intel Labs), Theodore Willke (Intel Labs)

HippogriffDB: Balancing I/O and GPU Bandwidth in Big Data Analytics

Jing Li (UCSD), Hung-Wei Tseng (UCSD), Chunbin Lin

(UCSD), Steven Swanson (UCSD), Yannis Papakonstantinou (UCSD)

Voodoo - A Vector Algebra for Portable Database Performance on Modern Hardware

Holger Pirk (MIT), Oscar Moll (MIT), Matei Zaharia (MIT), Samuel Madden (MIT)

Fast In-Memory SQL Analytics on Typed Graphs

Chunbin Lin (UCSD), Benjamin Mandel (UCSD), Yannis Papakonstantinou (UCSD), Matthias Springer (UCSD)

Recommendations and Skylines

Location: Auditorium 602

Chair: Hannes Mühlisen

Reconciling Skyline and Ranking Queries

Paolo Ciaccia (Università di Bologna), Davide Martinenghi (Politecnico di Milano)

Finding Diverse, High-Value Representatives on a Surface of Answers

You Wu (Google Research), Junyang Gao (Duke University), Pankaj Agarwal (Duke University), Jun Yang (Duke University)

An Experimental Evaluation of Point-of-interest Recommendation in Location-based Social Networks

Yiding Liu (Nanyang Technological University), Tuan-Anh Pham (Nanyang Technological University), Gao Cong (Nanyang Technological University), Quan Yuan (UIUC)

Heterogeneous Recommendations: What You Might Like To Read After Watching Interstellar

Rachid Guerraoui (EPFL), Anne-Marie Kermarrec (Inria), Tao Lin (EPFL), Rhicheck Patra (EPFL)

Approximation Structures

Location: Auditorium 606

Chair: Lefteris Sidiropoulos

Revisiting Reuse for Approximate Query Processing

Alex Galakatos (Brown University), Andrew Crotty (Brown University), Emanuel Zgraggen (Brown University), Carsten Binnig (Brown University), Tim Kraska (Brown University)

Probabilistic Database Summarization for Interactive Data Exploration

Laurel Orr (University of Washington), Dan Suciu (University of Washington), Magdalena Balazinska (University of Washington)

DigitHist: a Histogram-Based Data Summary with Tight Error Bounds

*Michael Shekelyan (Free University of Bozen-Bolzano),
Anton Dignös (Free University of Bozen-Bolzano), Johann
Gamper (Free University of Bozen-Bolzano)*

Data Driven Approximation with Bounded Resources

Yang Cao (University of Edinburgh), Wenfei Fan (University of Edinburgh)

Stream Processing 2

Location: Auditorium 601

Chair: Alan Fekete

Toward High-Performance Distributed Stream Processing via Approximate Fault Tolerance

Qun Huang (The Chinese Univ of Hong Kong), Patrick P. C. Lee (The Chinese Univ of Hong Kong)

ExtraV: Boosting Graph Processing Near Storage with a Coherent Accelerator (industrial)

Jinho Lee (IBM Research), Heesu Kim (Seoul National University), Sungjoo Yoo (Seoul National University), Kiyoung Choi (Seoul National University), Peter Hofstee (IBM Research), GiJoon Nam (IBM Research), Mark Nutter (IBM Research), Damir Jamsek (IBM Research)

Finding Persistent Items in Data Streams

Haipeng Dai (Nanjing University), Muhammad Shahzad (North Carolina State University), Alex X. Liu (Nanjing

University), Yuankun Zhong (Nanjing University)

Shrink - Prescribing Resiliency Solutions for Streaming

Badrish Chandramouli (Microsoft Research), Jonathan Goldstein (Microsoft Research)

Tutorial 4

Location: Auditorium 1601

New Trends on Exploratory Methods for Data Analytics

Davide Mottin (HPI), Matteo Lissandrini (University of Trento), Yannis Velegrakis (University of Trento), Themis Palpanas (Paris Descartes University)

Demo Group D

Location: Classrooms 2601, 2605, 2607

Flower: A Data Analytics Flow Elasticity Manager

Alireza Khoshkbarforoushha (Australian National University), Rajiv Ranjan (Newcastle University), Qing Wang (Australian National University), Carsten Friedrich (CSIRO)

LocLok: Location Cloaking with Differential Privacy via Hidden Markov Model

Yonghui Xiao (Emory University), Li Xiong (Emory University), Si Zhang (Jianghan University), Yang Cao

(Emory University)

MLog: Towards Declarative In-Database Machine Learning

Xupeng Li (Peking University), Bin Cui (Peking University), Yiru Chen (Peking University), Wentao Wu (Microsoft Research), Ce Zhang (ETH)

Foresight: Recommending Visual Insights

Çağatay Demiralp (IBM), Peter Haas (IBM), Srinivasan Parthasarathy (IBM), Tejaswini Pedapati (IBM)

ClaimBuster: The First-ever End-to-end Fact-checking System

Naeemul Hassan (University of Mississippi), Gensheng Zhang (University of Texas Arlington), Fatma Arslan (University of Texas Arlington), Josue Caraballo (University of Texas Arlington), Damian Jimenez (University of Texas Arlington), Siddhant Gawsane (University of Texas Arlington), Shohedul Hasan (University of Texas Arlington), Minumol Joseph (University of Texas Arlington), Aaditya Kulkarni (University of Texas Arlington), Anil Kumar Nayak (University of Texas Arlington), Vikas Sable (University of Texas Arlington), Chengkai Li (University of Texas at Arlington), Mark Tremayne (University of Texas Arlington)

QIRANA Demonstration: Real time Scalable Query Pricing

Shaleen Deep (UW-Madison), Paris Koutris (UW-Madison), Yash Bidasaria (UW-Madison)

Mind the Gap: Bridging Multi-Domain Query Workloads with EmptyHeaded

Christopher Aberger (Stanford University), Andrew Lamb (Stanford University), Kunle Olukotun (Stanford University), Christopher Ré (Stanford University)

Crossing the finish line faster when paddling the Data Lake with Kayak

Antonio Maccioni (Collective[i]), Riccardo Torlone (Roma Tre University)

QUIS: InSitu Heterogeneous Data Source Querying

Javad Chamanara (Friedrich Schiller University of Jena), Birgitta König-Ries (Friedrich Schiller University of Jena), H. V. Jagadish (University of Michigan)

Wednesday 08/30/2017 12:00-13:30

Lunch

Location: Tent

Wednesday 08/30/2017 13:30-15:00

Scalable Storage

Location: Audimax

Chair: Viktor Leis

Fast Scans on Key-Value Stores

Markus Pilman (ETH), Kevin Bocksrocker (Microsoft), Lucas Braun (ETH), Renato Marroquín (ETH), Donald Kossmann (ETH)

Caribou: Intelligent Distributed Storage

Zsolt Istvan (ETH), David Sidler (ETH), Gustavo Alonso (ETH)

The TileDB Array Data Storage Manager

Stavros Papadopoulos (Intel Labs and MIT), Kushal Datta (Intel Corporation), Samuel Madden (MIT), Timothy Mattson (Intel Labs)

PaxosStore: High-availability Storage Made Practical in WeChat (industrial)

Jianjun Zheng (Tencent Inc.), Qian Lin (NUS), Jiatao Xu (Tencent Inc.), Cheng Wei (Tencent Inc.), Chuwei Zeng (Tencent Inc.), Pingan Yang (Tencent Inc.), Yunfan Zhang (Tencent Inc.)

Crowdsourcing

Location: Auditorium 602

Chair: Li Xiong

DOCS: Domain-Aware Crowdsourcing System

Yudian Zheng (Hong Kong University), Guoliang Li (Tsinghua University), Reynold Cheng (Hong Kong University)

Understanding Workers, Developing Effective Tasks, and Enhancing Marketplace Dynamics: A Study of a Large Crowdsourcing Marketplace

Ayush Jain (University of Illinois), Akash Das Sarma (Stanford University), Aditya Parameswaran (UIUC), Jennifer Widom (Stanford University)

Truth Inference in Crowdsourcing: Is the Problem Solved?

Yudian Zheng (Hong Kong University), Guoliang Li (Tsinghua University), Yuanbing Li (Tsinghua University), Caihua Shan (Hong Kong University), Reynold Cheng (Hong Kong University)

A Data Quality Metric (DQM): How to Estimate the Number of Undetected Errors in Data Sets

Yeounoh Chung (Brown University), Sanjay Krishnan (UC Berkeley), Tim Kraska (Brown University)

Stream Processing 3

Location: Auditorium 606

Chair: Avrilia Floratou

History is a mirror to the future: Best-effort approximate complex event matching with insufficient resources

Zheng Li (Oracle), Tingjian Ge (University of Massachusetts Lowell)

Cohort Query Processing

Dawei Jiang (Zhejiang University), Qingchao Cai (NUS), Gang Chen (Zhejiang University), H. Jagadish (University of Michigan), Beng Chin Ooi (NUS), Kian-Lee Tan (NUS), Anthony Tung (NUS)

CarStream: An Industrial System of Big Data Processing for Internet-of-Vehicles (industrial)

Mingming Zhang (Beihang University), Tianyu Wo (Beihang University), Xuelian Lin (Beihang University), Tao Xie (University of Illinois), Yaxiao Liu (CAR Inc.)

A holistic view of stream partitioning costs

Nikos R. Katsipoulakis (University of Pittsburgh), Alexandros Labrinidis (University of Pittsburgh), Panos Chrysanthis (University of Pittsburgh)

Graphs and Networks 3

Location: Auditorium 601

Chair: Themis Palpanas

SkyGraph: Retrieving Regions of Interest using Skyline Subgraph Queries

Shiladitya Pande (IIT Madras), Sayan Ranu (IIT Delhi), Arnab Bhattacharya (IIT Kanpur)

Social Hash Partitioner: A Scalable Distributed Hypergraph Partitioner

Igor Kabiljo (Facebook), Brian Karrer (Facebook), Mayank

Pundir (Facebook), Sergey Pupyrev (Facebook), Alon Shalita (Facebook), Yaroslav Akhremtsev (Karlsruhe Institute of Technology), Alessandro Presta (Google)

READS: A Random Walk Approach for Efficient and Accurate Dynamic SimRank

minhao jiang (HKUST), Ada Wai Chee Fu (The CUHK), Raymond Chi-Wing Wong (HKUST), Ke Wang (Simon Fraser University)

Social Network Analysis

Location: Auditorium 670

Chair: Xiaofang Zhou

Resisting Tag Spam by Leveraging Implicit User Behaviors

Ennan Zhai (Yale University), Zhenhua Li (Tsinghua University), Zhenyu Li (Chinese Academy of Sciences), Fan Wu (Shanghai Jiaotong University), Guihai Chen (Shanghai Jiaotong University)

OLAK: An Efficient Algorithm to Prevent Unraveling in Social Networks

Fan Zhang (UTS), Wenjie Zhang (CSE), Ying Zhang (QCIS), Lu Qin (QCIS), Xuemin Lin (CSE)

Real-Time Influence Maximization on Dynamic Social Streams

Yanhao Wang (NUS), Qi Fan (NUS), Yuchen Li (NUS),

Kian-Lee Tan (NUS)

Revisiting the Stop-and-Stare Algorithms for Influence Maximization

Keke Huang (Nanyang Technological University), Sibo Wang (Nanyang Technological University), Glenn Bevilacqua (University of British Columbia), Xiaokui Xiao (Nanyang Technological University), Laks Lakshmanan (UBC)

Tutorial 5

Location: Auditorium 1601

Geometric Approaches for Top-k Queries

Kyriakos Mouratidis (Singapore Management University)

Demo Group C

Location: Classrooms 2601, 2605, 2607

C-Explorer: Browsing Communities in Large Graphs

Yixiang Fang (Hong Kong University), Reynold Cheng (Hong Kong University), Siqiang Luo (Hong Kong University), Jiafeng Hu (Hong Kong University), Kai Huang (Hong Kong University)

GRAPE: Parallelizing Sequential Graph Computations

Wenfei Fan (University of Edinburgh and Beihang Uni-

versity), Jingbo Xu (University of Edinburgh and Beihang University), Yinghui Wu (Washington State University), Wenyuan Yu (Beihang University), Jiaxin Jiang (Hong Kong Baptist University)

STEED: An Analytical Database System for TrEE-structured Data

Zhiyi Wang (Chinese Academy of Sciences), Dongyan Zhou (Chinese Academy of Sciences), Shimin Chen (Chinese Academy of Sciences)

Strider: An Adaptive, Inference-enabled Distributed RDF Stream Processing Engine

Xiangnan Ren (ATOS), Olivier Curé (UPEM LIGM - UMR CNRS 8049), Li Ke (ATOS), Jérémie Lhez (UPEM LIGM - UMR CNRS 8049), Badre Belabbess (ATOS), Tendry Randriamalala (ATOS), Yufan Zheng (ATOS), Gabriel Kepeklian (ATOS)

Explaining and Querying Knowledge Graphs by Relatedness

Valeria Fionda (University of Calabria), Giuseppe Pirrò (ICAR-CNR)

TeCoRe: Temporal Conflict Resolution in Knowledge Graphs

Melisachew Chekol (University of Mannheim), Giuseppe Pirrò (ICAR-CNR), Joerg Schoenfisch (University of Mannheim), Heiner Stuckenschmidt (University of

Mannheim)

PICASSO: Exploratory Search of Connected Subgraph Substructures in Graph Databases

Kai Huang (Fudan University), Sourav S Bhowmick (Nanyang Technological University), Shuigeng Zhou (Fudan University), Byron Choi (Hong Kong Baptist University)

Upsortable: Programming TopK Queries Over Data Streams

Julien Subercaze (Lab Hubert Curien), Christophe Gravier (Lab Hubert Curien), Syed Gillani (Lab Hubert Curien), Abderrahmen Kammoun (Lab Hubert Curien), Frédérique Laforest (Lab Hubert Curien)

Wednesday 08/30/2017 15:00-15:30

Coffee Break

Location: Tent

Wednesday 08/30/2017 15:30-17:00

Concurrency Control

Location: Audimax

Chair: Uwe Röhm

An Empirical Evaluation of In-Memory Multi-Version Concurrency Control

Yingjun Wu (NUS), Joy Arulraj (CMU), Jiexi Lin (CMU),

Ran Xian (CMU), Andrew Pavlo (CMU)

High Performance Transactions via Early Write Visibility

Jose Faleiro (Yale University), Daniel Abadi (Yale University), Joseph Hellerstein (UC Berkeley)

Mostly-Optimistic Concurrency Control for Highly Contended Dynamic Workloads on a Thousand Cores

Tianzheng Wang (University of Toronto), Hideaki Kimura (Hewlett Packard Enterprise)

Query Optimization

Location: Auditorium 602

Chair: Jens Teubner

Query Optimization for Dynamic Imputation

Jose Cambronero Sanchez (MIT), John Feser (MIT), Micah Smith (MIT), Samuel Madden (MIT)

Runtime Optimization of Join Location in Parallel Data Management Systems

Bikash Chandra (IIT Bombay), S. Sudarshan (IIT Bombay)

Non-Invasive Progressive Optimization for In-Memory Databases

Steffen Zeuch (Humboldt-Universität zu Berlin), Holger Pirk (MIT), Johann-Christoph Freytag (Humboldt-

Universität zu Berlin)

Looking Ahead Makes Query Plans Robust

Jianqiao Zhu (UW-Madison), Navneet Potti (UW-Madison), Saket Saurabh (UW-Madison), Jignesh Patel (UW-Madison)

Specialized Data Management

Location: Auditorium 606

Chair: Stefan Manegold

Comparative Evaluation of Big-Data Systems on Scientific Image Analytics Workloads

Parmita Mehta (University of Washington), Sven Dorkenwald (University of Washington), Dongfang Zhao (University of Washington), Tomer Kaftan (University of Washington), Alvin Cheung (University of Washington), Magdalena Balazinska (University of Washington), Ariel Rokem (University of Washington), Andrew Connolly (University of Washington), Jacob Vanderplas (University of Washington), Yusra AlSayyad (University of Washington)

NG-DBSCAN: Scalable Density-Based Clustering for Arbitrary Data

Alessandro Lulli (University of Pisa), Matteo Dell'Amico (Symantec Research Labs), Pietro Michiardi (EURECOM), Laura Ricci (University of Pisa)

Optimizing Deep CNN-Based Queries over Video Streams at Scale

Daniel Kang (Stanford University), John Emmons (Stanford University), Firas Abuzaid (Stanford University), Peter Bailis (Stanford University), Matei Zaharia (Stanford University)

Fast and Adaptive Indexing of Multi-Dimensional Observational Data

Sheng Wang (NUS), David Maier (Portland State University), Beng Chin Ooi (NUS)

Graphs and Networks 4

Location: Auditorium 601

Chair: Marco Serafini

Truss-based Community Search: a Truss-equivalence Based Indexing Approach

Esra Akbas (Florida State University), Peixiang Zhao (Florida State University)

Efficient Computation of Feedback Arc Set at Web-Scale

Michael Simpson (University of Victoria), Venkatesh Srinivasan (University of Victoria), Alex Thomo (University of Victoria)

Effective Community Search over Large Spatial Graphs

Yixiang Fang (Hong Kong University), Reynold Cheng (Hong Kong University), Xiaodong Li (Hong Kong University), Siqiang Luo (Hong Kong University), Jiafeng Hu (Hong Kong University)

From Community Detection to Community Profiling

Hongyun Cai (ADSC), Vincent Zheng (Advanced Digital Sciences Cent), Fanwei Zhu (Zhejiang University City College), Kevin Chen-Chuan Chang (UIUC), Zi Huang (The University of Queensland)

Data Mining and Analytics

Location: Auditorium 670

Chair: Essam Mansour

Revenue Maximization in Incentivized Social Advertising

Cigdem Aslay (ISI Foundation), Francesco Bonchi (ISI Foundation), Laks Lakshmanan (UBC), Wei Lu (LinkedIn)

Interactive Time Series Exploration Powered by the Marriage of Similarity Distances

Rodica Neamtu (Worcester Polytechnic Institute), Ramoza Ahsan (Worcester Polytechnic Institute), Elke Rundensteiner (Worcester Polytechnic Institute), Gabor Sarkozy (Worcester Polytechnic Institute)

MapReduce and Streaming Algorithms for Diver-

sity Maximization in Metric Spaces of Bounded Doubling Dimension

Matteo Ceccarello (University of Padova), Andrea Pietraccaprina (University of Padova), Geppino Pucci (University of Padova), Eli Upfal (Brown University)

Local Search Methods for k-Means with Outliers

Shalmoli Gupta (UIUC), Ravi Kumar (Google), Kefu Lu (Washington University), Benjamin Moseley (Washington University St. Louis), Sergei Vassilvitskii (Google)

Tutorial 6

Location: Auditorium 1601

Spatial Crowdsourcing: Challenges, Techniques, and Applications

Yongxin Tong (Beihang University), Lei Chen (HKUST), Cyrus Shahabi (USC)

Demo Group A

Location: Classrooms 2601, 2605, 2607

A Confidence-Aware Top-k Query Processing Toolkit on Crowdsourcing

Yan Li (University of Macau), Ngai Meng Kou (University of Macau), Hao Wang (Nanjing University), Leong Hou U (University of Macau), Zhiguo Gong (University of Macau)

DataTweener: A Demonstration of a Tweening Engine for Incremental Visualization of Data Transforms

Meraj Ahmed Khan (Ohio State University), Larry Xu (UC Berkeley), Arnab Nandi (Ohio State University), Joseph Hellerstein (UC Berkeley)

A Demonstration of Stella: A Crowdsourcing-Based Geotagging Framework

Christopher Jonathan (University of Minnesota), Mohamed Mokbel (University of Minnesota)

Interactive Navigation of Open Data Linkages

Erkang Zhu (University of Toronto), Ken Pu (UOIT), Fateh-meh Nargesian (University of Toronto), Renee Miller (University of Toronto)

noWorkflow: a Tool for Collecting, Analyzing, and Managing Provenance from Python Scripts

João Felipe Pimentel (Universidade Federal Fluminense), Leonardo Murta (Universidade Federal Fluminense), Vanessa Braganholo (Universidade Federal Fluminense), Juliana Freire (NYU)

ARShop: A Cloud-based Augmented Reality System for Shopping

Chao Wang (NUS), Yihao Feng (Dartmouth College), Qi Guo (NUS), Zhaoxian Li (NUS), Kexin Liu (NUS), Zijian Tang (NUS), Anthony Tung (NUS), Lifu Wu (NUS), Yuxin

Zheng (NUS)

Debugging Transactions and Tracking their Provenance with Reenactment

Xing Niu (IIT), Bahareh Sadat Arab (Illinois Institute of Technology), Seokki Lee (Illinois Institute of Technology), Su Feng (Illinois Institute of Technology), Xun Zou (Illinois Institute of Technology), Dieter Gawlick (Oracle), Vasudha Krishnaswamy (Oracle), Zhen Hua Liu (Oracle), Boris Glavic (Illinois Institute of Technology)

FlashView: An Interactive Visual Explorer for Raw Data

Zhifei Pang (Zhejiang University), Sai Wu (Zhejiang University), Gang Chen (Zhejiang University), Ke Chen (Zhejiang University), Lidan Shou (Zhejiang University)

Automating Data Citation in CiteDB

Abdussalam Alawini (University of Pennsylvania), Susan Davidson (University of Pennsylvania), Wei Hu (University of Pennsylvania), Yinjun Wu (University of Pennsylvania)

Wednesday 08/30/2017 17:00-18:00

Poster Reception

Location: Audimax Foyer

Poster Reception

Poster presentations for all papers that were presented that same day, accompanied by some drinks. Additionally, two VLDB Journal papers will present a poster.

Avoiding class warfare: Managing Continuous Queries with Differentiated Classes of Service

Thao N. Pham (University of Pittsburgh), Panos K. Chrysanthis (University of Pittsburgh), Alexandros Labrinidis (University of Pittsburgh)

PANDA: Towards Partial Topology-based Search on Large Networks in a Single Machine

Miao Xie (Nanyang Technological University and Chinese Academy of Sciences and Huawei), Sourav S. Bhowmick (Nanyang Technological University), Gao Cong (Nanyang Technological University), Qing Wang (Chinese Academy of Sciences)

Wednesday 08/30/2017 19:00-23:00

“VLDB Octoberfest” Banquet

Location: Hofbräuhaus

Thursday 08/31/2017 08:30-10:30

Endowment Awards + Demo Award + Jens Dittrich plenary

Location: Audimax

Women in Database Research Award Talk: 7 Se-

crets That My Mother Didn't Tell Me

Tova Milo (Tel Aviv University)

What does it take to be a good researcher? And, is it different when you are a woman? These are questions that many of us are wondering about throughout our career. Being honored with a VLDB Women in Database Research Award, I would like to share with you in this talk some of the secrets to successful research that I have learned over the years. These secrets highlight some of the fundamental research directions that I have taken. No less importantly, they explain how I successfully got to work on them, both personally and professionally.

Early Career Award Talk: Human-in-the-loop Data Integration

Guoliang Li (Tsinghua University)

Data integration aims to integrate data in different sources and provide users with a unified view. However, data integration cannot be completely addressed by purely automated methods. We propose a hybrid human-machine data integration framework that harnesses human ability to address this problem, and apply it initially to the problem of entity matching. The framework first uses rule-based algorithms to identify possible matching pairs and then utilizes the crowd to compute actual matching pairs from these candidate pairs. In the first step, we propose similarity-based

rules and knowledge-based rules to obtain the candidate matching pairs, develop effective algorithms to learn these rules based on positive and negative examples, and build a distributed in-memory system to efficiently apply these rules. In the second step, we propose a selection-inference-refine framework that uses the crowd to verify the candidate pairs. We first select some representative tasks to ask the crowd, use transitivity rules and partial order to infer the answers of unasked tasks based on the crowd results of the asked tasks, and refine the inferred answers with low confidence to improve the quality. We develop a crowd-powered database system CDB that allows users to utilize a SQL-like language for processing crowd-based queries. Finally, we provide emerging challenges in human-in-the-loop data integration.

Ten Year Best Paper Award Talk: Intelligent Probing for Locality Sensitive Hashing: Multi-Probe LSH and Beyond

Qin Lv (University of Colorado Boulder), William Josephson (Solano Labs), Zhe Wang (Datrium), Moses Charikar (Stanford University), Kai Li (Princeton University)

The past decade has been marked by the (continued) explosion of diverse data content and the fast development of intelligent data analytics techniques. One problem we identified in the mid-2000s was similarity search

of feature-rich data. The challenge here was achieving both high accuracy and high efficiency in high-dimensional spaces. Locality sensitive hashing (LSH), which uses certain random space partitions and hash table lookups to find approximate nearest neighbors, was a promising approach with theoretical guarantees. But LSH alone was insufficient since a large number of hash tables were required to achieve good search quality. Building on an idea of Panigrahy, our multi-probe LSH method introduces the idea of intelligent probing. Given a query object, we strategically probe its neighboring hash buckets (in a query-dependent fashion) by calculating the statistical probabilities of similar objects falling into each bucket. Such intelligent probing can significantly reduce the number of hash tables while achieving high quality. In this paper, we revisit the problem motivation, the challenges, the key design considerations of multi-probe LSH, as well as discuss recent developments in this space and some questions for further research.

Deep Learning (m)eats Databases

Jens Dittrich (Saarland University)

Imagine a machine that is able to compose music and write poems; paint realistic artificial images and dream up video from textual descriptions; paint pictures or entire videos in the style of any artist; translate in-

between any pair of natural languages. A machine that can recognize any content in images and videos; diagnose diseases, imitate spoken language – in any voice. A machine that wins games thought to be exclusive to human intelligence. All of that with superhuman performance of course. Sounds like science fiction? Well, then welcome to the year 2017! Currently we are witnessing the biggest revolution in computer science since the invention of the Internet. Deep Learning is shaking the world of computer science and overrunning entire (sub-)disciplines. In this talk I will briefly sketch some of the recent advances in deep learning and what they have to do with databases. Where are synergies? Where should we be looking at? This talk will have a particular focus on recent technical developments in the intersection of databases and/or deep learning in Europe.

Jens Dittrich is a Full Professor of Computer Science in the area of Databases, Data Management, and Big Data at Saarland University, Germany. Previous affiliations include U Marburg, SAP AG, and ETH. He received an Outrageous Ideas and Vision Paper Award at CIDR 2011, a BMBF VIP Grant in 2011, a best paper award at VLDB 2014 (the second ever given to an E&A paper), two CS teaching awards in 2011 and 2013, as well as several presentation awards including a qualification

for the interdisciplinary German science slam finals in 2012 and three presentation awards at CIDR (2011, 2013, and 2015). He has been a PC member and area chair/group leader of prestigious international database conferences and journals such as PVLDB/VLDB, SIGMOD, ICDE, and VLDB Journal. At Saarland University he co-organizes the Data Science Summer School (<http://datasciencemaster.de>).

Since 2013 he has been teaching some of his classes on data management as flipped classrooms. See <http://datenbankenlernen.de> or <http://youtube.com/jensdit> for a list of freely available videos on database technology in German (introduction to databases) and English (database architectures and implementation techniques). He is also author of a “flipped textbook” on databases. Since 2016 he has been working on a start-up at the intersection of deep learning and databases.

His research focuses on fast access to big data including in particular: data analytics on large datasets, main-memory databases, database indexing, reproducability, and deep learning.

Thursday 08/31/2017 10:30-10:55

Coffee Break

Location: Tent

Thursday 08/31/2017 10:55-12:00

Transactions and Persistence

Location: Audimax

Chair: Shimin Chen

Write-Behind Logging

Joy Arulraj (CMU), Matthew Perron (CMU), Andrew Pavlo (CMU)

PHyTM: Persistent Hybrid Transactional Memory

Hillel Avni (Huawei), Trevor Brown (University of Toronto)

Parallel Replication across Formats in SAP HANA for Scaling Out Mixed OLTP/OLAP Workloads (industrial)

Juchang Lee (SAP Labs Korea), SeungHyun Moon (POSTECH), Kyu Hwan Kim (SAP Labs Korea), Deok Hoe Kim (SAP Labs Korea), Sang Kyun Cha (Seoul National University), Wook-Shin Han (POSTECH), Chang Gyoo Park (SAP Labs Korea), Hyoung Jun Na (SAP Labs Korea), Joo Yeon Lee (SAP Labs Korea)

Data Access

Location: Auditorium 602

Chair: Torsten Grust

Slalom: Coasting Through Raw Data via Adaptive Partitioning and Indexing

Matthaios Olma (EPFL), Manos Karpathiotakis (EPFL), Ioannis Alagiannis (Microsoft), Manos Athanassoulis (Harvard University), Anastasia Ailamaki (EPFL)

A Forward Scan based Plane Sweep Algorithm for Parallel Interval Joins

Panagiotis Bouros (Aarhus University), Nikos Mamoulis (Hong Kong University)

Skipping-oriented Partitioning for Columnar Layouts

Liwen Sun (UC Berkeley), Michael Franklin (UC Berkeley), Jiannan Wang (UC Berkeley), Eugene Wu (Columbia University)

Data Statistics

Location: Auditorium 606

Chair: Raymond Ng

Statisticum: Data Statistics Management in SAP HANA (industrial)

Anisoara Nica (SAP), Reza Sherkat (SAP), Mihnea Andrei (SAP), Xun Chen (SAP), Martin Heidel (SAP), Christian Bensberg (SAP), Heiko Gerwens (SAP)

Bias-Aware Sketches

Jiecao Chen (Indiana University), Qin Zhang (Indiana

University)

Adaptive Statistics in Oracle 12c (industrial)

Mohamed Zait (Oracle), Sunil Chakkappen (Oracle), Suratna Budalakoti (Oracle Labs), Satyanarayana Valluri (Oracle), Ramarajan Krishnamachari (Oracle), Alan Wood (Oracle Labs)

Potpourri

Location: Auditorium 601

Chair: Chengkai Li

Data Vocalization: Optimizing Voice Output of Relational Data

Immanuel Trummer (Cornell University), Jiancheng Zhu (Cornell University), Mark Bryan (Cornell University)

Perturbation Analysis of Database Queries

Brett Walenz (Duke University), Jun Yang (Duke University)

Dscaler: Synthetically Scaling A Given Relational Database

Jiangwei Zhang (NUS), Y.C. Tay (NUS)

Thursday 08/31/2017 12:00-13:30

Lunch

Location: Tent

Thursday 08/31/2017 13:30-15:00

Data Partitioning

Location: Audimax

Chair: Khuzaima Daudjee

AdaptDB: Adaptive Partitioning for Distributed Joins

Yi Lu (MIT), Anil Shanbhag (MIT), Alekh Jindal (Microsoft), Samuel Madden (MIT)

Adaptive NUMA-aware data placement and task scheduling for analytical workloads in main-memory column-stores

Iraklis Psaroudakis (EPFL), Tobias Scheuer (SAP SE), Norman May (SAP SE), Abdelkader Sellami (SAP SE), Anastasia Ailamaki (EPFL)

Clay: Fine-Grained Adaptive Partitioning for General Database Schemas

Marco Serafini (Qatar Computing Research Institute), Rebecca Taft (MIT), Aaron J. Elmore (University of Chicago), Andrew Pavlo (CMU), Ashraf Aboulnaga (Qatar Computing Research Institute), Michael Stonebraker (MIT)

An Experimental Comparison of Partitioning Strategies in Distributed Graph Processing

Shiv Verma (University Of Illinois at Urbana-Champaign), Luke Leslie (UIUC), Yosub Shin (Samsara), Indranil Gupta

Graphs and Networks 5

Location: Auditorium 602

Chair: Sourav S Bhowmick

HubPPR: Effective Indexing for Approximate Personalized PageRank

Sibo Wang (Nanyang Technological University), Youze Tang (Nanyang Technological University), Xiaokui Xiao (Nanyang Technological University), Yin Yang (Hamad Bin Khalifa University), Zengxiang Li (Institute of High Performance Computing)

A General Framework for Estimating Graphlet Statistics via Random Walk

Xiaowei Chen (CUHK), Yongkun Li (University of Science and Technology of China), Pinghui Wang (Xi'an Jiaotong University), John C.S. Lui (The CUHK)

An Experimental Evaluation of SimRank-based Similarity Search Algorithms

Zhipeng Zhang (Peking University), Yingxia Shao (PKU), Bin Cui (Peking University), Ce Zhang (ETH)

On Sampling from Massive Graph Streams

Nesreen Ahmed (Intel Labs), Nick Duffield (Texas A&M University), Theodore Willke (Intel Labs), Ryan Rossi (PARC)

Visualization

Location: Auditorium 606

Chair: Guoliang Li

I've Seen "Enough": Incrementally Improving Visualizations to Support Rapid Decision Making

Sajjadur Rahman (UIUC), Maryam Aliakbarpour (MIT), Ha Kyung Kong (UIUC), Eric Blais (University of Waterloo), Karrie Karahalios (UIUC), Aditya Parameswaran (UIUC), Ronitt Rubinfeld (MIT)

ASAP: Prioritizing Attention via Time Series Smoothing

Kexin Rong (Stanford University), Peter Bailis (Stanford University)

Effortless Data Exploration with zenvisable: An Expressive and Interactive Visual Analytics System

Tarique Ashraf Siddiqui (UIUC), Albert Kim (MIT), John Lee (UIUC), Karrie Karahalios (UIUC), Aditya Parameswaran (UIUC)

Data Tweening: Incremental Visualization of Data Transforms

Meraj Ahmed Khan (The Ohio State University), Larry Xu (UC Berkeley), Arnab Nandi (Ohio State University), Joseph Hellerstein (UC Berkeley)

Location: Auditorium 601

Chair: Henrik Muehe

LFTF: A Framework for Efficient Tensor Analytics at Scale

Fan Yang (CUHK), Fanhua Shang (CUHK), Yuzhen Huang (CUHK), James Cheng (CUHK), Jinfeng Li (CUHK), Yunjian Zhao (CUHK), Ruihao Zhao (CUHK)

Dimensions Based Data Clustering and Zone Maps (industrial)

Mohamed Ziauddin (Oracle), Andrew Witkowski (Oracle), You Jung Kim (Oracle), Janaki Lahorani (Oracle), Dmitry Potapov (Oracle), Murali Krishna (Oracle)

Scalable Asynchronous Gradient Descent Optimization for Out-of-Core Models

Chengjie Qin (UC Merced), Martin Torres (UC Merced), Florin Rusu (University of California)

Fiber-based architecture for NFV cloud databases (industrial)

Vaidas Gasiunas (Huawei), David Dominguez-Sal (Huawei), Ralph Acker (Huawei), Aharon Avitzur (Huawei), Ilan Bronshtein (Huawei), Rushan Chen (), Eli Ginot (Huawei), Norbert Martinez (Huawei), Michael Müller (Huawei), Alexander Nozdrin (Huawei), Weijie Ou

*(Huawei), Nir Pachter (Huawei), Dima Sivov (Huawei),
Eliezer Levy (Huawei)*

Estimation and Approximation

Location: Auditorium 670

Chair: Johann Gamper

Pyramid Sketch: a Sketch Framework for Frequency Estimation of Data Streams

Tong Yang (Peking University), Yang Zhou (Peking University), Hao Jin (Peking University), Shigang Chen (University of Florida), Xiaoming Li (Peking University)

Sapprox: Enabling Efficient and Accurate Approximations on Sub-datasets with Distribution-aware Online Sampling

Xuhong Zhang (University of Central Florida), Jun Wang (University of Central Florida), Jiangling Yin (University of Central Florida), Shouling Ji (Georgia Institute of Technology)

Estimating Quantiles from the Union of Historical and Streaming Data

Sneha Singh (Iowa State University), Divesh Srivastava (AT&T), Srikanta Tirthapura (Iowa State University)

Effective and Complete Discovery of Order Dependencies via Set-based Axiomatization

Jaroslaw Szlichta (UOIT), Parke Godfrey (York University)

sity), Lukasz Golab (University of Waterloo), Mehdi Kar-gar (University of Windsor), Divesh Srivastava (AT&T)

Tutorial 7

Location: Auditorium 1601

The Era of Big Spatial Data

Ahmed Eldawy (UC Riverside), Mohamed Mokbel (Uni-versity of Minnesota)

Demo Group D

Location: Classrooms 2601, 2605, 2607

Flower: A Data Analytics Flow Elasticity Manager

Alireza Khoshkbarforoushha (Australian National Uni-versity), Rajiv Ranjan (Newcastle University), Qing Wang (Australian National University), Carsten Friedrich (CSIRO)

LocLok: Location Cloaking with Differential Pri-vacy via Hidden Markov Model

Yonghui Xiao (Emory University), Li Xiong (Emory Uni-versity), Si Zhang (Jianghan University), Yang Cao (Emory University)

MLog: Towards Declarative In-Database Machine Learning

Xupeng Li (Peking University), Bin Cui (Peking Univer-sity), Yiru Chen (Peking University), Wentao Wu (Mi-

crosoft Research), Ce Zhang (ETH)

Foresight: Recommending Visual Insights

Çağatay Demiralp (IBM), Peter Haas (IBM), Srinivasan Parthasarathy (IBM), Tejaswini Pedapati (IBM)

ClaimBuster: The First-ever End-to-end Fact-checking System

Naeemul Hassan (University of Mississippi), Gensheng Zhang (University of Texas Arlington), Fatma Arslan (University of Texas Arlington), Josue Caraballo (University of Texas Arlington), Damian Jimenez (University of Texas Arlington), Siddhant Gawsane (University of Texas Arlington), Shohedul Hasan (University of Texas Arlington), Minumol Joseph (University of Texas Arlington), Aaditya Kulkarni (University of Texas Arlington), Anil Kumar Nayak (University of Texas Arlington), Vikas Sable (University of Texas Arlington), Chengkai Li (University of Texas at Arlington), Mark Tremayne (University of Texas Arlington)

QIRANA Demonstration: Real time Scalable Query Pricing

Shaleen Deep (UW-Madison), Paris Koutris (UW-Madison), Yash Bidasaria (UW-Madison)

Mind the Gap: Bridging Multi-Domain Query Workloads with EmptyHeaded

Christopher Aberger (Stanford University), Andrew Lamb

(Stanford University), Kunle Olukotun (Stanford University), Christopher Ré (Stanford University)

Crossing the finish line faster when paddling the Data Lake with Kayak

Antonio Maccioni (Collective[i]), Riccardo Torlone (Roma Tre University)

QUIS: InSitu Heterogeneous Data Source Querying

Javad Chamanara (Friedrich Schiller University of Jena), Birgitta König-Ries (Friedrich Schiller University of Jena), H. V. Jagadish (University of Michigan)

Thursday 08/31/2017 15:00-15:30

Coffee Break

Location: Tent

Thursday 08/31/2017 15:30-17:00

DB Engines 2

Location: Audimax

Chair: Allison Holloway

Two Birds, One Stone: A Fast, yet Lightweight, Indexing Scheme for Modern Database Systems

Jia Yu (Arizona State University), Mohamed Sarwat (Arizona State University)

SAP HANA Adoption of Non-Volatile Memory (industrial)

Mihnea Andrei (SAP), Christian Lemke (SAP), Günter Radestock (SAP), Robert Schulze (SAP), Carsten Thiel (SAP), Rolando Blanco (SAP), Akanksha Meghlan (SAP), Muhammad Sharique (SAP), Sebastian Seifert (SAP), Surendra Vishnoi (SAP), Daniel Booss (SAP SE), Thomas Peh (SAP), Ivan Schreter (SAP), Werner Thesing (SAP), Mehul Wagle (SAP), Thomas Willhalm (Intel Deutschland GmbH)

Memory Management Techniques for Large-Scale Persistent-Main-Memory Systems

Ismail Oukid (TU Dresden & SAP SE), Daniel Booss (SAP SE), Adrien Lespinasse (Independent), Wolfgang Lehner (TU Dresden), Thomas Willhalm (Intel Deutschland GmbH), Grégoire Gomes (Grenoble INP - Ensimeag)

Query Processing

Location: Auditorium 602

Chair: Ioana Manolescu

SquirrelJoin: Network-Aware Distributed Join Processing with Lazy Partitioning

Lukas Rupprecht (Imperial College London), William Culhane (Imperial College London), Peter Pietzuch (Imperial College London)

Reverse Engineering Aggregation Queries

Wei Chit Tan (SUTD), Meihui Zhang (SUTD), Hazem Elmeleegy (Turn Inc.), Divesh Srivastava (AT&T)

Leveraging Set Relations in Exact Set Similarity Join

Xubo Wang (CSE), Lu Qin (QCIS), Xuemin Lin (CSE), Ying Zhang (QCIS), Lijun Chang (CSE)

Auto-Join: Joining Tables by Leveraging Transformations

Erkang Zhu (University of Toronto), Yeye He (Microsoft Research), Surajit Chaudhuri (Microsoft Research)

Text and Semi-Structured

Location: Auditorium 606

Chair: Azza Abouzied

Knowledge Exploration using Tables on the Web

Fernando Chirigati (NYU), Jialu Liu (University of Illinois at Urbana-Champaign), Flip Korn (Google Research), You Wu (Google Research), Cong Yu (Google), Hao Zhang (Google Research)

KBQA: Learning Question Answering over QA Corpora and Knowledge Bases

Wanyun Cui (Fudan University), Yanghua Xiao (Fudan University), Haixun Wang (Facebook), Yangqiu Song (West Virginia University), Seung-won Hwang (Yonsei

University), Wei Wang (Fudan University)

Provenance for Natural Language Queries

Daniel Deutch (Tel Aviv University), Nave Frost (Tel Aviv University), Amir Gilad (Tel Aviv University)

SilkMoth: An Efficient Method for Finding Related Sets with Maximum Matching Constraints

Dong Deng (MIT), Albert Kim (MIT), Samuel Madden (MIT), Michael Stonebraker (MIT)

Spatial Data Management 3

Location: Auditorium 601

Chair: Jianliang Xu

Minimal OnRoad Time Route Scheduling on Time-Dependent Graph

Lei Li (University of Queensland), Wen Hua (University of Queensland), Xingzhong Du (University of Queensland), Xiaofang Zhou (University of Queensland)

A General and Parallel Platform for Mining Co-Movement Patterns over Large-scale Trajectories

Qi Fan (NUS), Dongxiang Zhang (NUS), Huayu Wu (NUS), Kian-Lee Tan (NUS)

Dimensional Testing for Reverse k-Nearest Neighbor Search

Guillaume Casanova (ONERA-DCSD), Elias Englmeier

(LMU), Michael Houle (NII), Peer Kroeger (LMU), Michael Nett (Google), Erich Schubert (LMU), Arthur Zimek (SDU)

Pivot-based Metric Indexing

Lu Chen (Zhejiang University), Yunjun Gao (Zhejiang University), Baihua Zheng (Singapore Management University), Christian Jensen (Aalborg University), Hanyu Yang (Zhejiang University), Keyu Yang (Zhejiang University)

Distributed Systems and Cloud 2

Location: Auditorium 670

Chair: Jeff Pound

Lifting the Haze off the Cloud: A Consumer-Centric Market for Database Computation in the Cloud

Yue Wang (University of Massachusetts Amherst), Alexandra Meliou (University of Massachusetts Amherst), Jerome Miklau (University of Massachusetts Amherst)

Query-able Kafka: An agile data analytics pipeline for mobile wireless networks (industrial)

Eric Falk (University of Luxembourg), Vijay Gurbani (Bell Labs), Radu State (University of Luxembourg)

Price-Optimal Querying with Data APIs

Prasang Upadhyaya (University of Washington), Mag-

dalena Balazinska (University of Washington), Dan Suciu (University of Washington)

Cümülön-D: Data Analytics in a Dynamic Spot Market

Botong Huang (Duke University), Jun Yang (Duke University)

Tutorial 8

Location: Auditorium 1601

Complex Event Recognition in the Big Data Era

Nikos Giatrakos (Technical University of Crete), Alexander Artikis (NCSR Demokritos), Antonios Deligiannakis (Technical University of Crete), Minos Garofalakis (Technical University of Crete)

Demo Group B

Location: Classrooms 2601, 2605, 2607

Thoth in Action: Memory Management in Modern Data Analytics

Mayuresh Kunjir (Duke University), Shivnath Babu (Duke University)

Monopedia: Staying Single is Good Enough - The HyPer Way for Web Scale Applications

Maximilian Schüle (TUM), Pascal Schliski (TUM), Thomas Hutzelmann (TUM), Tobias Rosenberger (TUM), Viktor

*Leis (TUM), Dimitri Vorona (TUM), Alfons Kemper (TUM),
Thomas Neumann (TUM)*

Dima: A Distributed In-Memory Similarity-Based Query Processing System

Ji Sun (Tsinghua University), Zeyuan Shang (Tsinghua University), Guoliang Li (Tsinghua University), Dong Deng (MIT), Zhifeng Bao (RMIT University)

A BAD Demonstration: Towards Big Active Data

Steven Jacobs (University of California at Riverside), Md Yusuf Sarwar Uddin (University of California at Irvine), Michael Carey (University of California at Irvine), Vagelis Hristidis (University of California at Riverside), Vassilis Tsotras (University of California at Riverside), Nalini Venkatasubram (University of California at Irvine), Yao Wu (University of California at Irvine), Syed Safir (University of California at Irvine), Purvi Kaul (University of California at Irvine), Xikui Wang (University of California at Irvine), Mohiuddin Abdul Qader (University of California), Yawei Li (University of California at Riverside)

A Demonstration of ST-Hadoop: A MapReduce Framework for Big Spatio-temporal Data

Louai Alarabi (University of Minnesota), Mohamed Mokbel (University of Minnesota)

Creation and Interaction with Large-scale

Domain-Specific Knowledge Bases

Shreyas Bharadwaj (IBM Watson), Laura Chiticariu (IBM Research-Almaden), Marina Danilevsky (IBM Research-Almaden), Samarth Dhingra (IBM Watson), Samved Divekar (IBM Watson), Arnaldo Carreno-Fuentes (IBM Watson), Himanshu Gupta (IBM Research-India), Nitin Gupta (IBM Research-India), Sang-Don Han (IBM Watson), Mauricio Hernandez (IBM Research-Almaden), Howard Ho (IBM Watson), Parag Jain (IBM Research-India), Salil Joshi (IBM Research-India), Hima Karanam (IBM Research-India), Saravanan Krishnan (IBM Research-India), Rajasekar Krishnamurthy (IBM Research-Almaden), Yunyao Li (IBM Research-Almaden), Satishkumar Manivannan (IBM Watson), Ashish Mittal (IBM Research-India), Fatma Ozcan (IBM Research-Almaden), Abdul Quamar (IBM Research-Almaden), Poornima Raman (IBM Watson), Diptikalyan Saha (IBM Research-India), Karthik Sankaranarayanan (IBM Research-India), Jaydeep Sen (IBM Research-India), Prithviraj Sen (IBM Research-Almaden), Shivakumar Vaithyanathan (IBM Watson), Mitesh Vasa (IBM Watson), Hao Wang (IBM Watson), Huaiyu Zhu (IBM Research-Almaden)

Exploring big volume sensor data with Vroom

Oscar Moll (MIT), Aaron Zalewski (MIT), Sudeep Pillai (MIT), Samuel Madden (MIT), Michael Stonebraker (MIT),

Vijay Gadepally (MIT Lincoln Labs)

DITIR: Distributed Index for High Throughput Trajectory Insertion and Real-time Temporal Range Query

Ruichu Cai (Guangdong University of Technology), Zijie Lu (Guangdong University of Technology), Li Wang (Advanced Digital Sciences Center), Zhenjie Zhang (Advanced Digital Sciences Center), Tom Fu (Advanced Digital Sciences Center), Marianne Winslett (University of Illinois at Urbana-Champaign)

Thursday 08/31/2017 17:00-18:00

Poster Reception

Location: Audimax Foyer

Poster Reception

Poster presentations for all papers that were presented that same day, accompanied by some drinks. Additionally, two VLDB Journal papers will present a poster.

Geo-Social Group Queries with Minimum Acquaintance Constraints

Qijun Zhu (Hong Kong Baptist University), Haibo Hu (Hong Kong Polytechnic University), Cheng Xu (Hong Kong Baptist University), Jianliang Xu (Hong Kong Baptist University), Wang-Chien Lee (Pennsylvania State University)

Argument Discovery via Crowdsourcing

*Quoc Viet Hung Nguyen (University of Queensland),
Chi Thang Duong (EPFL), Thanh Tam Nguyen (EPFL),
Matthias Weidlich (Humboldt-Universität zu Berlin), Karl
Aberer (EPFL), Hongzhi Yin (University of Queensland),
Xiaofang Zhou (University of Queensland and Macau
University of Science and Technology)*

Workshops

Monday 08/27/2017 08:30-17:00

FADS

Location: Auditorium 602

Failed Aspirations in Database Systems

Spyros Blanas (Ohio State University), Justin Lewandoski (Microsoft Research), Andy Pavlo (CMU)

BIRTE

Location: Auditorium 606

Eleventh International Workshop on Real-Time Business Intelligence and Analytics

Malu Castellanos (Teradata Aster), Panos K Chrysanthis (University of Pittsburgh)

TPCTC

Location: Auditorium 601

Ninth TPC Technology Conference on Performance Evaluation & Benchmarking

Raghunath Niambar (Cisco), Meikel Poess (Oracle)

VLIoT

Location: Auditorium 670

The International Workshop on Very Large Internet of Things

Sven Groppe (University of Lübeck), Carlo Alberto Boano (Graz University of Technology)

PhD Workshop

Location: Auditorium 1601

VLDB PhD Workshop

Peter Christen (The Australian National University), Bettina Kemme (McGill University), Erhard Rahm (University of Leipzig)

Friday 09/01/2017 08:30-17:00

MATES

Location: Auditorium 602

Workshop on Mobility Analytics for Spatio-temporal and Social Data

Christos Doulkeridis (University of Piraeus), Qiang Qu (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences)

ADMS

Location: Auditorium 606

Eight International Workshop on Accelerating Analytics and Data Management Systems Using Modern Processor and Storage Architectures

Rajesh Bordawekar (IBM Watson), Tirthankar Lahiri (Oracle)

DMAH

Location: Auditorium 601

The Third International Workshop on Data Management and Analytics for Medicine and Healthcare

Fusheng Wang (Stony Brook University), Gang Luo (University of Washington), Edmon Begoli (Oak Ridge National Laboratory)

DBPL

Location: Auditorium 670

The Sixteenth International Symposium on Database Programming Languages

Tiark Rompf (Purdue University), Alexander Alexandrov (TU Berlin)

BOSS

Location: Auditorium 1601

Third Workshop on Big Data Open Source Systems

Tyson Condie (UCLA), Tilmann Rabl (TU Berlin)

THE WORLD'S FIRST ENGAGEMENT DATABASE

Today's customers aren't just shoppers. They're seekers. That means they spend a lot more time interacting than transacting. **And you have only 8 seconds to grab their attention** before they start looking somewhere else.

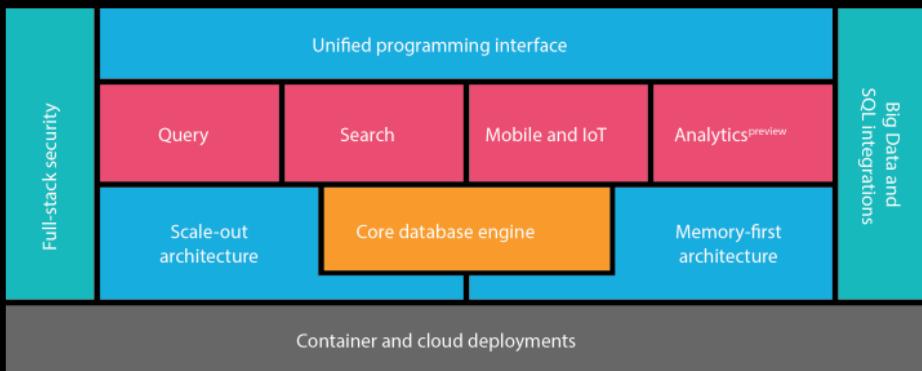
The Couchbase Data Platform makes every interaction count.

To make the most of every moment, you need an **Engagement Database** that delivers an amazing customer experience (CX) every time.

HERE'S WHY

- 1 By 2020, CX will overtake price and product as the key brand differentiator
- 2 Strong CX companies will retain 89% of customers versus 33% for weak CX companies
- 3 70% of Global 500 brands already have dedicated digital experience teams in place

MEET THE COUCHBASE DATA PLATFORM



Start creating amazing customer experiences today.

1-650-417-7500 | www.couchbase.com



Software Engineer, Infrastructure

Google

Software Engineering



Google's software engineers develop the next-generation technologies that change how billions of users connect, explore, and interact with information and one another. Our products need to handle information at massive scale, and extend well beyond web search. We're looking for engineers who bring fresh ideas from all areas, including information retrieval, distributed computing, large-scale system design, networking and data storage, security, artificial intelligence, natural language processing, UI design and mobile; the list goes on and is growing every day. As a software engineer, you will work on a specific project critical to Google's needs with opportunities to switch teams and projects as you and our fast-paced business grow and evolve. We need our engineers to be versatile, display leadership qualities and be enthusiastic to tackle new problems across the full-stack as we continue to push technology forward.

As a Software Engineer working on Google's infrastructure, you have the opportunity to work on everything from the core platform that runs the world's largest distributed network to redefining the systems that allow applications and services to provide useful information to billions of users around the globe. From our Data Center software groups to Google's Cloud Platform, Gmail to YouTube, our infrastructure engineers across departments wrestle with the vast scale of a ubiquitous system, its products, and services and revolutionize industry leading technologies to handle the sheer magnitude at which Google operates.

Google is and always will be an engineering company. We hire people with a broad set of technical skills who are ready to tackle some of technology's greatest challenges and make an impact on millions, if not billions, of users. At Google, engineers not only revolutionize search, they routinely work on massive scalability and storage solutions, large-scale applications and entirely new platforms for developers around the world. From AdWords to Chrome, Android to YouTube, Social to Local, Google engineers are changing the world one technological achievement after another.

Responsibilities

- Build our platforms, systems and infrastructure using your strong background in distributed systems and large scale storage systems.
- Manage individual projects priorities, deadlines and deliverables with your technical expertise.
- Design, develop, test, deploy, maintain, and enhance software solutions.

Qualifications

Minimum qualifications:

- BA/BS degree in Computer Science or related technical field or equivalent practical experience.
- 4 years of relevant work experience, including software development experience, or 1 year of relevant work experience with a PhD in Computer Science or related technical field.
- Professional coding experience in C/C++, Java, Python or Go.
- Experience architecting and developing large scale distributed systems. Experience in concurrency, multithreading and synchronization.

Preferred qualifications:

- MS or PhD in Computer Science.
- Experience with Unix/Linux environments.
- Experience with TCP/IP and network programming.
- Experience with database internals, database language theories, database design, SQL and database programming.
- Understanding of technologies such as virtualization and global infrastructure, load balancing, networking, massive data storage, Hadoop, MapReduce and security.
- Interest or exposure to networking technologies/concepts such as Software Defined Networking (SDN) and OpenFlow.

Microsoft Research

ADVANCE YOUR CAREER



CONNECT WITH US



BOOST YOUR RESEARCH



STAY INFORMED



```
100110010110  
1010010100  
10000110011  
1010110010  
1111011100  
1011011100
```



Together, we can achieve more

At Microsoft Research, we're inventing the future of computing. We relentlessly push the boundaries of technology, actively collaborate with world-class researchers, and passionately support the next generation of scientists.

Engage with us: Microsoft.com/research

Push a Button

Move Your Database to the Oracle Cloud



... or Back to Your Data Center

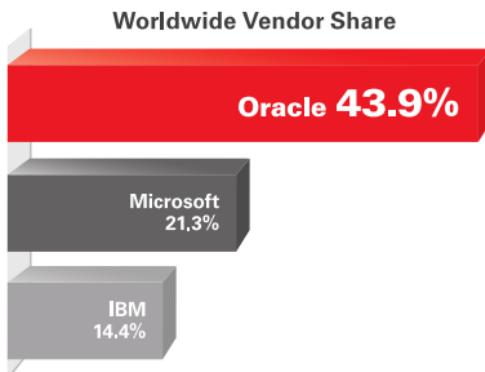
ORACLE®

cloud.oracle.com/database

STILL

#1

Database



Oracle Database

Trusted by 310,000 Customers Worldwide

ORACLE®

oracle.com/database

Source: IDC, "Worldwide Relational Database Management Systems Software Market Shares, 2015: The Year of Transition to the Cloud," IDC #US41484516, June 2016; Table 1 (Worldwide Relational Database Management Systems Revenue by Vendor). Vendor share based on software license and maintenance revenue. Copyright © 2017, Oracle and/or its affiliates. All rights reserved.

Enriching lifestyles with Information Technology

Recruit Institute of Technology (RIT) is the technology hub and research lab for Recruit Holdings, a company that provides over 200 online services in the areas of human resources, travel, housing, education, restaurants and many other areas in which people make daily lifestyle decisions.

We conduct research in several areas, including data management, data integration, natural language processing, machine learning, and artificial intelligence. We collaborate with universities and publish in top-notch conferences.



Example Project: BigGorilla



BigGorilla is an open-source data integration and data preparation ecosystem (currently in Python) to enable data scientists to perform integration and analysis of data. BigGorilla brings decades of research on data integration into an open-source platform with the goal of accelerating progress in the field and adoption of its techniques. <http://www.biggorilla.org>





At Tableau, innovation is key.

We constantly push the boundaries of Visual Analytics. Tableau is expanding its engineering and R&D power to continue to delight customers with the best analytics platform ever built. You can be part of the team that delivers these breakthroughs.

We are currently hiring:

- Software Engineers in Development
- Software Engineers in Test
- Research Engineers

Visit careers.tableau.com for all open positions.

Haven't had a chance to check out Tableau?

Students & Professors download your free trial today at
www.tableau.com/academic

See yourself at



Jewel Loree

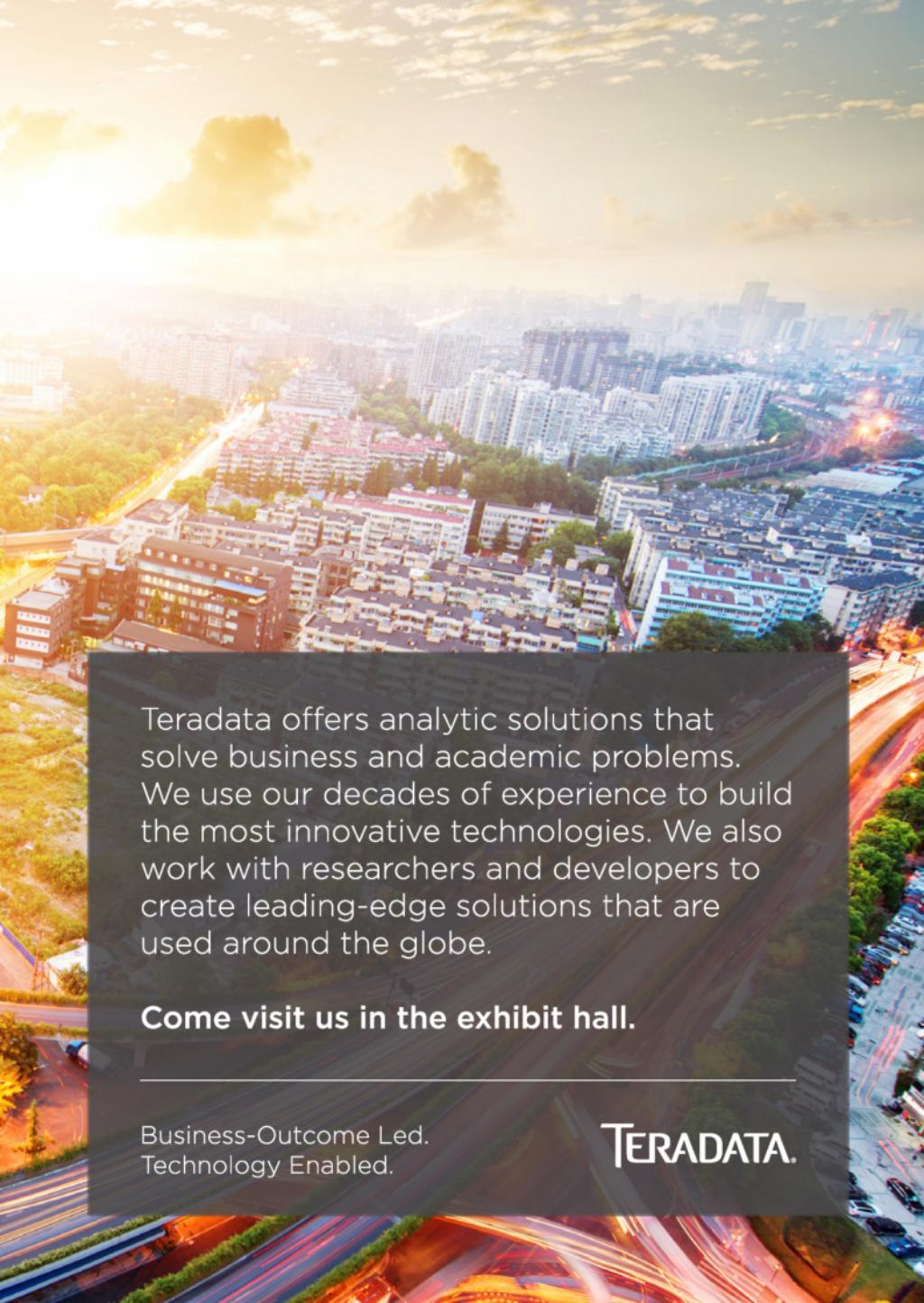
Product Manager, Visual Analytics

Bassist, Indie/Surf band



Advanced Analytics at Enterprise Scale





Teradata offers analytic solutions that solve business and academic problems. We use our decades of experience to build the most innovative technologies. We also work with researchers and developers to create leading-edge solutions that are used around the globe.

Come visit us in the exhibit hall.

Business-Outcome Led.
Technology Enabled.

TERADATA

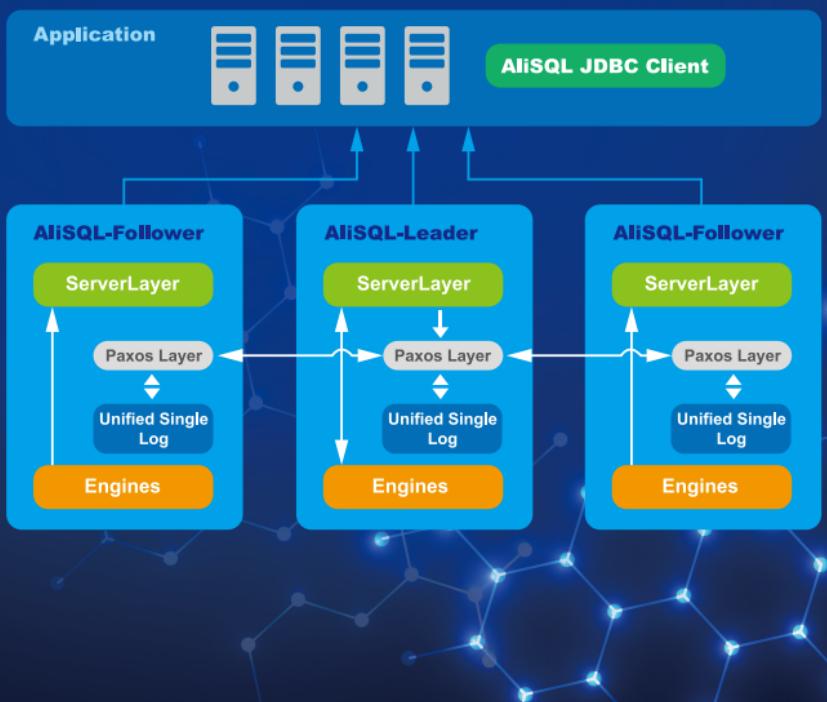
Alibaba Cloud ApsaraDB

ApsaraDB is a Database-as-a-Service platform provided by Alibaba Cloud.

It covers the mainstream database engines including the most popular open-source databases like MySQL (AliSQL), PostgreSQL and Redis, as well as commercial databases like SQL Server and PPAS. It also provides the hybrid analysis database HybridDB and off-line analysis database E-MapReduce, which are compatible with MySQL and PostgreSQL protocols.

AliSQL

AliSQL is a MySQL branch maintained by Alibaba Database team, which supports all lines of business within Alibaba group including Alibaba Cloud, Taobao, Tmall and Ant Financial. We introduced lots of features to AliSQL to support dramatic business growth. Besides providing a fantastic shopping experience in Taobao/Tmall double 11 festival, AliSQL also has more than 100, 000 running RDS instances and 50, 000 users in the cloud. To maximize the outcome we can get from MySQL, we also implemented a high performance Paxos protocol. With the help of Paxos, we can deploy our AliSQL Cluster geo-distributed. Our vision is quite clear that we want to introduce the world the fastest OLTP database at the lowest cost.



Connecting the World

The world's largest social graph

Building the tools & systems to help 1.9 billion people across the world connect, communicate and share requires constant innovation. At Facebook, research permeates everything we do.

The Facebook platform is our lab for research, development, and innovation. Our talented teams of researchers and engineers are constant innovators as they design and build the next generation systems to serve the 1.9 billion people who use our products.

We're hiring!

We're looking for talented industry and academic researchers to join our team!

Visit us at research.fb.com/careers to learn more about a career at Facebook.



Database Systems Expert

Location: Munich Germany

Huawei is a leading global information and communications technology solutions provider.

The European Research Institute (ERI) performs strategic research and cutting-edge development for Huawei.

The Databases group in ERI is advancing database technology for future use in telecommunication and enterprise and is looking for a few technical experts that will take part in that awarding endeavor. It operates out of sites in

Germany and Israel.

What you can expect:

- Research towards conception followed by architecture, design, prototyping, and development.
- Analysis and evaluation of relevant state of the art in the academia and industry, and competitors' products.
- Performing business analysis and technical risk evaluation of proposed designs and plans.

Interested? You are invited to contact the recruiting manager directly

Eliezer Levy, tel. +972 54 2277128
eliezer.levy@huawei.com

HUAWEI TECHNOLOGIES
Duesseldorf GmbH
German Research Center, Munich
Riesstraße 25
80992 Munich, Germany



your future made with IBM

IBM Research is hiring

We live in a moment of remarkable change and opportunity. Data and technology are transforming industries and societies, ushering in a new era of Cognitive Computing. IBM Research is a leader in this worldwide transformation, building on a long history of innovation.

For more than seven decades, IBM Research has defined the future of technology. Our scientists, among them six Nobel Laureates and six Turing Award winners, have produced ten U.S. National Medals of Technology and five U.S. National Medals of Science. Along the way we helped put a man on the moon, defeated Kasparov at chess, and built a *Jeopardy!*[®] champion named Watson.

At IBM, you can achieve what others think is impossible. And in doing so, you'll play a significant role in shaping the future. Join us and discover what you can make of this moment.

What will you make with IBM?

ibm.com/jobs



Join the SAP HANA Database Campus!

Are you passionate about Software Development and working on your IT degree? Do you want to start your career by working on the next generation database technologies?

Then join us – The SAP HANA Database Campus team!

The SAP HANA team develops a platform that performs parallel in-memory processing of huge data sets to offer extremely fast real-time responses for analytic and transactional queries. The platform also provides libraries for predictive, planning, text processing, spatial, and business analytics – all on the same architecture.

Learn more about the SAP HANA Database Campus here:



hana.sap.com

AEROSPIKE

THE HIGH PERFORMANCE, NOSQL
DATABASE FOR REAL-TIME, MISSION
CRITICAL APPLICATIONS.

- * Predictable performance
- * Highest up time & availability
- * Lowest Total Cost of Ownership

www.aerospike.com



EXASOL

The world's fastest*
in-memory database for
large-scale data analytics.

*For over 3 years, we are still unbeaten in the TPC-H
benchmarks. No-one else comes close. Period.

Get started with EXASOL today
wwwexasolcom/vldb



We are fearless in our innovation.

THINK WHAT WE CAN DO FOR YOUR CAREER.

We're seeking great minds like you to help us build the next generation in database technology.

Join the team that is helping the world's most sophisticated organizations transform their industries by harnessing the power of data.



www.MongoDB.com/Careers



@MongoDB



www.Facebook.com/MongoDB



@MongoDB + @MongoDBCareers



Persistent Systems
Delivering the Blueprint
for a Software Driven Business

Join the team that obsessively focuses on the "**how**" of digital.

www.persistent.com