

Online Program

**Australasian Database Conference
2015**

June 4-7, 2015
Conference Room 1 & 2
Storey Hall
RMIT University, Melbourne

Time	Thu 4 June PhD School	Fri 5 June PhD School	Sat 6 June PhD School & Conference Day 1	Sun 7 June Conference Day 2
8:30	Registration Opens		Opening Address & Best Paper Awards	
9:00	Lu Qin Graph Processing in the Era of Big Data	Gerhard Weikum Knowledge Graphs: from a Fistful of Triples to Deep Data and Deep Text	Bingsheng He Emerging HPC Technologies for Real-Time (Big) Data Analytics: A Tutorial	Research Session 3: Data Analytics and Mining
10:30	Morning Tea	Morning Tea	Morning Tea	Morning Tea
11:00	Tamer Ozsü An Overview of Graph Data Management and Analysis	Gerhard Weikum Knowledge Graphs: from a Fistful of Triples to Deep Data and Deep Text	Bingsheng He Emerging HPC Technologies for Real-Time (Big) Data Analytics: A Tutorial	Research Session 4: Information Retrieval and Integration
12:30	Lunch	Lunch	Lunch (Demo)	Lunch
14:00	Tamer Ozsü An Overview of Graph Data Management and Analysis	Cyrus Shahabi Spatial Indexing and Spatial Crowdsourcing of User-Generated- Video	Research Session 1: Spatiotemporal Databases	Research Session 5: Graphs and Social Networks
15:30	Afternoon Tea	Afternoon Tea	Afternoon Tea	
16:00	Panel Discussion for ECR/PhD: How to develop career and apply for grants	Posters Session	Research Session 2: Query Processing and Optimization	
	17.30 Reception		18.30 Banquet	17:30 Close

Graph Processing in the Era of Big Data

Lu Qin

University of Technology, Sydney

Abstract. With the emergence and rapid proliferation of applications that deal with big graphs, such as web graphs (Google, Yahoo), social networks (Facebook, Twitter), e-commerce networks (Amazon, Ebay), and road networks, graph processing has become increasingly prevalent and important in recent years. However, in the era of big data, the explosion and profusion of available graph data in a wide range of application domains rise up new challenges and opportunities in graph processing. In this talk, I will first investigate these new challenges. To tickle these challenges, I will then introduce our recent research on big graph processing in terms of new graph query semantics, new query processing algorithms, new graph indexing techniques, and new computing paradigms. Finally, I will discuss our potential future research directions for graph processing.

Short Biography. Dr. Lu Qin received his PhD degree in 2010 from the department of Systems Engineering and Engineering Management (SEEM) in the Chinese University of Hong Kong (CUHK). Dr. Qin is currently a core member in the Centre of Quantum Computation and Intelligent Systems (QCIS) at the University of Technology, Sydney (UTS). Dr. Qin's research interests include algorithm design and analysis on big data, big graph processing in the cloud, and big graph searching and mining. He has published 50+ top-tier conference/journal papers including 7 SIGMOD papers, 9 VLDB papers, and 7 ICDE papers in the top-3 database conferences, and 7 VLDB journal papers, 1 Algorithmica paper, and 2 TKDE papers in the top-ranked database and algorithm journals. His book entitled "Keyword Search in Databases" is the first monograph on keyword search in databases. Dr. Qin served as a program committee member of a lot of top database and data mining conferences. He has received several research funds from Australian government and UTS.

An Overview of Graph Data Management and Analysis

M. Tamer Özsu

University of Waterloo

Abstract. Graphs have always been important data types for database researchers. With the recent growth of social networks, Wikipedia, Linked Data, RDF, and other networks, the interest in managing very large graphs have again gained momentum. In this talk I will first present a taxonomy of graph processing systems and then summarize research on querying and analytics over property graphs and management and querying of RDF graphs.

Short Biography. M. Tamer Özsu is Professor of Computer Science at the David R. Cheriton School of Computer Science, and Associate Dean (Research) of the Faculty of Mathematics at the University of Waterloo. His research is in data management focusing on large-scale data distribution and management of non-traditional data. He is a Fellow of the Association for Computing Machinery (ACM), and of the Institute of Electrical and Electronics Engineers (IEEE), an elected member of the Science Academy of Turkey, and member of Sigma Xi and American Association for the Advancement of Science (AAAS). He currently holds a Cheriton Faculty Fellowship at the University of Waterloo.

Knowledge Graphs: from a Fistful of Triples to Deep Data and Deep Text

Gerhard Weikum

Max Planck Institute for Informatics

Abstract. Knowledge graphs (KG's), aka. knowledge bases, are huge repositories of entities, their types, properties, and relationships between entities. KG's have become a key asset for search, analytics, recommendations, and data integration on the Web and in enterprises. Rooted in academic research and community projects such as DBpedia, Freebase, and Yago, KG's are now intensively used at big industrial stakeholders such as Google, Microsoft, Yahoo, Alibaba, Bloomberg, Walmart, and many others.

This talk reviews the knowledge graph technology, discussing strengths and limitations and pointing out opportunities for further research. The talk spans a spectrum of issues that arise in the life-cycle and use-cases of a KG: construction from data and text sources, maintaining over time, extension with common sense knowledge, querying and mining, boosting language understanding and text analytics, and usability issues in interactive exploration.

Short Biography. Gerhard Weikum is a Scientific Director at the Max Planck Institute for Informatics in Saarbruecken, Germany, and also an Adjunct Professor at Saarland University. He graduated from the University of Darmstadt, Germany.

Weikum's research spans transactional and distributed systems, self-tuning database systems, DB&IR integration, and the automatic construction of knowledge bases from Web and text sources. He co-authored a comprehensive textbook on transactional systems, received the VLDB 10-Year Award for his work on automatic DB tuning, and is one of the creators of the YAGO knowledge base.

Gerhard Weikum is an ACM Fellow, a member of the Academia Europaea, and a member of several academies in Germany. He has served on various editorial boards, including Communications of the ACM, ACM TODS and ACM TWEB, and as PC chair of conferences like ACM SIGMOD, Data Engineering, and CIDR. From 2003 through 2009 he was president of the VLDB Endowment. He received a Google Focused Research Award in 2010, the ACM SIGMOD Contributions Award in 2011, and an ERC Synergy Grant in 2013.

Emerging HPC Technologies for Real-Time (Big) Data Analytics:

A Tutorial

Bingsheng He

Nanyang Technological University

Abstract. Big data has become a buzz word. Among various big-data challenges, real-time data analytics has been identified as one of the most exciting and promising areas for both academia and industry. We are facing the challenges at all levels ranging from sophisticated algorithms and procedures to mine the gold from massive data to high-performance computing (HPC) techniques and systems to get the useful data in time. In this tutorial, we review the system design and implementation of HPC technologies (including GPUs, FPGAs and RDMA etc) as weapons to address the performance requirement of real-time data analytics. Particularly, we focus on the interplay between HPC and real-time data analytics, where real-time data analytics also poses significant challenges to the design and implementation of HPC technologies. I will also present our recent research efforts in developing real-time data analytics systems by Xtra Computing Group at NTU Singapore (<http://pdcc.ntu.edu.sg/xtra/>) as well as related research from other groups. Finally, I will outline some open problems in this field.

Short Biography. Dr. Bingsheng He is currently an Assistant Professor at Division of Networks and Distributed Systems, School of Computer Engineering, Nanyang Technological University. Before that, he held a research position in the System Research group of Microsoft Research Asia (2008 - 2010), where his major research was building high performance cloud computing systems for Microsoft. He got the Bachelor degree in Shanghai Jiao Tong University (1999 - 2003), and the Ph.D. degree in Hong Kong University of Science & Technology (2003 - 2008). His current research interests include cloud computing, database systems and high performance computing. His papers are published in prestigious international journals (such as ACM TODS, IEEE TKDE/TPDS/TC) and proceedings (such as ACM SIGMOD, VLDB/PVLDB, ACM/IEEE SuperComputing, PACT, HPDC, ACM SoCC, and CIDR). He has been awarded with the IBM Ph.D. fellowship (2007 - 2008) and with NVIDIA Academic Partnership (2010 - 2011).

Spatial Indexing and Spatial Crowdsourcing of User-Generated-Video

Cyrus Shahabi

University of Southern California

Abstract. I will start by showing a demo of our MediaQ system prototype. MediaQ is a novel online media management system to collect, organize, share, and search user-generated mobile videos (UGV) from the public. Subsequently, I focus on two of the research challenges underlying MediaQ. First, I discuss our approach to index and search UGVs more effectively by utilizing the smartphone sensors (e.g., GPS locations, compass directions) to geo-tag each video frame by the spatial extent of its coverage area. Next, I introduce our generic framework for spatial crowdsourcing and discuss various techniques for optimal assignment of spatiotemporal tasks (e.g., UGV data collection) to human workers. Finally, I conclude by summarizing our ongoing efforts in spatial crowdsourcing.

Short Biography. Cyrus Shahabi is a Professor of Computer Science and Electrical Engineering and the Director of the Information Laboratory (InfoLAB) at the Computer Science Department and also the Director of the NSF's Integrated Media Systems Center (IMSC) at the University of Southern California (USC). He is also the director of Informatics at USC's Viterbi School of Engineering. He was the CTO and co-founder of a USC spin-off, Geosemble Technologies, which was acquired in July 2012. Since then, he founded another company, ClearPath, focusing on predictive path-planning for car navigation systems. He received his B.S. in Computer Engineering from Sharif University of Technology in 1989 and then his M.S. and Ph.D. Degrees in Computer Science from the University of Southern California in May 1993 and August 1996, respectively. He authored two books and more than two hundred research papers in the areas of databases, GIS and multimedia with an h-index of 43. He also holds more than 12 US Patents.

Dr. Shahabi is a fellow of IEEE, and a recipient of the ACM Distinguished Scientist award in 2009, the 2003 U.S. Presidential Early Career Awards for Scientists and Engineers (PECASE), the NSF CAREER award in 2002, and the 2001 Okawa Foundation Research Grant for Information and Telecommunications.

Research Session 1: Spatiotemporal Databases

Unifying Spatial, Temporal and Semantic Features for an Effective GPS Trajectory-Based Location Recommendation

Hamidu Abdel-Fatao, Jiuyong Li, Jixue Liu

Storing and Processing Massive Trajectory Data on SAP HANA

Haozhou Wang, Kai Zheng, Hoyoung Jeung, Shane Bracher, Asadul Islam, Wasim Sadiq, Shazia Sadiq, Xiaofang Zhou

Bus Arrival Time using a Modified Amalgamation of Fuzzy Clustering and Neural Network on Spatio-temporal Data

Sonia Khetarpaul, Shyam Kumar Gupta, Shikhar Malhotra, L. Venkata Subramaniam

TK-SK: Textual-restricted K Spatial Keyword Query on Road Networks

Xiaopeng Kuang, Pengpeng Zhao, Victor S. Sheng, Jian Wu, Zhixu Li, Guanfeng Liu, Zhiming Cui

Effective Spatial Keyword Query Processing on Road Networks

Hailin Fang, Pengpeng Zhao, Victor S. Sheng, Jian Wu, Jiajie Xu, An Liu, Zhiming Cui

Research Session 2: Query Processing and Optimization

Handling Query Skew in Large Indexes: A View Based Approach

Weihuang Huang, Jeffrey Xu Yu, Zechao Shang

Using Word Embeddings to Enhance Keyword Identification for Scientific Publications

Rui Wang, Wei Liu, Chris McDonald

Presto-RDF: SPARQL Querying over Big RDF Data

Mulugeta Mammo, Srividya K. Bansal

Improvement of Join Algorithms for Low-Selectivity Joins on MapReduce

Akiyoshi MATONO, Hirotaka OGAWA, Isao KOJIMA

A Cache-based Semi-Stream Join to deal with Unmatched Stream Data

M. Asif Naeem, Imran Sarwar Bajwa, Noreen Jamil

Research Session 3: Data Analytics and Mining

Predicting Passengers in Public Transportation Using Smart Card Data

Mengyu Dou, Tieke He, Hongzhi Yin, Xiaofang Zhou, Bin Luo, Zhenyu Chen

Predicting Users' Purchasing Behaviors Using Their Browsing History

Tieke He, Hongzhi Yin, Zhenyu Chen, Bin Luo, Xiaofang Zhou

Efficient Discovery of Differential Dependencies through Association Rules Mining

Selasi Kwashie, Jixue Liu, Jiuyong Li, Feiyue Ye

Efficient Mining of Non-derivable Emerging Patterns

Vincent Mwintieru Nofong, Jixue Liu, Jiuyong Li

Truth discovery in Material Science databases

Eve Belisle, Zi Huang, Aimen Gheribi

Research Session 4: Information Retrieval and Integration

Ontology Augmentation via Attribute Extraction from Multiple Types of Sources

Xiu Susie Fang, Xianzhi Wang, Quan Z. Sheng

Community Based Information Dissemination

Zhengwei Yang, Ada Wai-Chee Fu, Yanyan Xu, Silu Huang, Ho Fung Leung

A Domain Independent Approach for Extracting Terms from Research Papers

Birong Jiang, Endong Xun, Jianzhong Qi

Cognition and Statistical-based Crowd Evaluation Framework for ER-In-house Crowdsourcing System: Inbound Contact Center

Morteza Saberi, Omar Khadeer Hussain, Naeem Khalid Janjua, Elizabeth Chang

A Fast and Effective Image Geometric Verification Method for Efficient CBIR

Ling-Bo Kong, Ling-Hai Kong, Tao Yang, Wei Lu

Research Session 5: Graphs and Social Networks

Belief Revision in Uncertain Data Integration

Fereidoon Sadri

Predicting the Spread of a New Tweet in Twitter

Md Musfique Anwar, Jianxin Li, Chengfei Liu

Personal Process Description Graph for Describing and Querying Personal Processes

Jing Xu, Hye-young Paik, Anne H. H. Ngu, Liming Zhan

Detecting Spamming Groups in Social Media Based on Latent Graph

Qunyan Zhang, Chi Zhang, Peng Cai, Weining Qian, Aoying Zhou

Demo Papers (Lunch, Saturday 06-June-2015)

A Framework of Enriching Business Processes Life-Cycle with Tagging Information

Zakaria Maamar, Sherif Sakr, Ahmed Barnawi, Seyed-Mehdi-Reza Behehsti

SocialTrail: Recommending Social Trajectories from Location-Based Social Networks

Qinzhe Zhang, Litao Yu, Guodong Long

SocialAnalysis: A Real-time Query and Mining System from Social Media Data Streams

Haishuai Wang, Peng Zhang, Ling Chen, Chengqi Zhang

SCIT: A Schema Change Interpretation Tool for Dynamic-Schema Data Warehouses

Rihan Hai, Vasileios Theodorou, Maik Thiele, Wolfgang Lehner

SEMI: A Scalable Entity Matching System based on MapReduce

Pingfu Chao, Yuming Li, Zhu Gao, Junhua Fang, Xiaofeng He, Rong Zhang