

NDBC2016
NATIONAL DATABASE CONFERENCE

第33届中国数据库学术会议

中国，深圳 · 2016年10月14日-16日



深圳大学



会场

会议地点：广东省深圳市南山区白石路深圳大学南校区计算机楼1楼会议厅。

WIFI接入点名称：SZU_WLAN, NDBC2016_1, NDBC2016_2, 无需密码。



交通

因2016年全国大众创业万众创新活动周及第二届国际创客周实行交通管制，班车和步行时间差不多，步行线路经过双创周展区，鼓励前往。

日期	时间	方向	乘车地点
10月13日	17: 30	NDBC会场 → 晚餐地点	计算机与软件学院楼下
	晚餐后（预计19: 30）	晚餐地点 → 后海酒店	深圳大学教工餐厅
10月14日	07: 50	后海酒店 → NDBC会场	凯宾斯基酒店/新桃园酒店/凯利达酒店
	晚餐后（预计19: 30）	晚餐地点 → 后海酒店	前海渔港
	22: 00	NDBC会场 → 后海酒店	计算机与软件学院楼下
10月15日	07: 40	后海酒店 → NDBC会场	凯宾斯基酒店/新桃园酒店/凯利达酒店
	08: 30-11: 30	班车（每半小时）：后海酒店 → NDBC会场	凯宾斯基酒店/新桃园酒店/凯利达酒店
	13: 30-18: 00		
	09: 30-12: 00	班车（每半小时）：NDBC会场 → 后海酒店	计算机与软件学院楼下
	13: 30-18: 00		
晚餐后（预计21: 30）	晚餐地点 → 后海酒店	前海渔港	
10月16日	08: 20	后海酒店 → NDBC会场	凯宾斯基酒店/新桃园酒店/凯利达酒店
	09: 00-11: 30	班车（每半小时）：后海酒店 → NDBC会场	凯宾斯基酒店/新桃园酒店/凯利达酒店
	13: 30-17: 30		
	10: 00-12: 00	班车（每半小时）：NDBC会场 → 后海酒店	计算机与软件学院楼下
	13: 30-17: 30		
晚餐后（预计19: 30）	晚餐地点 → 后海酒店	深圳大学教工餐厅	

乘车须知：

1. 每次发车前10分钟，酒店大堂有学生引导到达乘车点。
2. 酒店所在街区道路狭窄，无法长时间停车，请参会人员准时达到酒店大堂，过时不候。
3. 联系人王海涛；电话 186 8159 7915。

会议组织

大会名誉主席 Honorary Chair

陈国良（深圳大学）

程序委员会主席 PC Chairs

李国良（清华大学）

毛睿（深圳大学）

彭智勇（武汉大学）

出版主席 Publication Chairs

徐建良（香港浸会大学）

张滇（深圳大学）

组织主席 Local Chairs

林佳利（深圳大学）

许红龙（佛山科学技术学院）

林子雨（厦门大学）

李斌阳（香港中文大学）

文集主席 Proceedings Chair

吴定明（深圳大学）

大会主席 General Chairs

李清泉（深圳大学）

于戈（东北大学）

黄锦辉（香港中文大学）

宣传主席 Publicity Chairs

范举（人民大学）

李荣华（深圳大学）

演示主席 Demo Chairs

陈跃国（人民大学）

伍楷舜（深圳大学）

会议网站主席 Website Chair

林佳利（深圳大学）

财务主席 Treasure Chairs

黄国顺（佛山科学技术学院）

刘刚（深圳大学）

主办单位： 中国计算机协会
China Computer Federation

承办单位： 中国计算机学会数据库专业委员会
China Computer Federation Technical Committee on Databases

 深圳大学
SHENZHEN UNIVERSITY

协办单位： 厦门大学

 香港中文大学
CUHK Faculty of Engineering

 佛山科学技术学院
FOSHAN UNIVERSITY

 华智集团(深圳)有限公司

赞助单位： HUAWEI

 人大金仓
Kingbase

 iiBD
Institutional Center for Industrial Big Data
北京工业大数据研究中心

 高等教育出版社
HIGHER EDUCATION PRESS

 电子工业出版社
ELECTRONIC INDUSTRY PRESS

 人民邮电出版社
POSTS & TELECOM PRESS

 红亚科技
HONGYAA

 清华大学出版社
TSINGHUA UNIVERSITY PRESS

 睿业训
RUIYEXUN

 卓维大数据
ZHUOWEI DATA

审稿人

柏文阳 (南京大学)
陈恩红 (中国科学技术大学)
陈 红 (中国人民大学)
陈世敏 (中国科学院)
崔 斌 (北京大学)
党德鹏 (北京师范大学)
段 磊 (四川大学)
冯 铃 (清华大学)
高云君 (浙江大学)
郭景峰 (燕山大学)
黄锦辉 (香港中文大学)
金澈清 (华东师范大学)
景 宁 (国防科学技术大学)
李 川 (四川大学)
李国徽 (华中科技大学)
李红燕 (北京大学)
李庆忠 (山东大学)
李荣华 (深圳大学)
李石君 (武汉大学)
廖国琼 (江西财经大学)
刘红岩 (清华大学)
刘玉葆 (中山大学)
毛 睿 (深圳大学)
潘海为 (哈尔滨工程大学)
钱铁云 (武汉大学)
秦小麟 (南京航空航天大学)
万常选 (江西财经大学)
汪 卫 (复旦大学)
王 晨 (清华大学)
王丽珍 (云南大学)
王建勇 (清华大学)
王晓玲 (华东师范大学)
王意洁 (国防科技大学)
商 烁 (中国石油大学)
宋宝燕 (辽宁大学)
孙焕良 (沈阳建筑大学)
孙未未 (复旦大学)
汤 庸 (中山大学)
徐建良 (香港浸会大学)
徐立臻 (东南大学)
杨晓春 (东北大学)
印 鉴 (中山大学)
禹晓辉 (山东大学)
袁晓洁 (南开大学)
张东祥 (电子科技大学)
张 岩 (北京大学)
周皓峰 (IBM中国软件开发中心)
邹兆年 (哈尔滨工业大学)

蔡 毅 (华南理工大学)
陈 刚 (浙江大学)
陈 群 (西北工业大学)
陈晓云 (兰州大学)
崔立真 (山东大学)
杜小勇 (中国人民大学)
范 举 (中国人民大学)
高 军 (北京大学)
关佶红 (同济大学)
黄健斌 (西安电子科技大学)
黄冬梅 (上海海洋大学)
金培权 (中国科技大学)
李爱萍 (太原理工大学)
李翠平 (中国人民大学)
李国良 (清华大学)
李金宝 (黑龙江大学)
李陶深 (广西大学)
李盛恩 (山东建筑大学)
李舟军 (北京航空航天大学)
刘国华 (燕山大学)
刘奇志 (南京大学)
马 帅 (北京航空航天大学)
牛保宁 (太原理工大学)
彭成宝 (东软集团)
钱卫宁 (华东师范大学)
乔少杰 (西南交通大学)
汪传建 (石河子大学)
王朝坤 (清华大学)
王宏志 (哈尔滨工业大学)
王建民 (清华大学)
王腾蛟 (北京大学)
王 鑫 (天津大学)
任家东 (燕山大学)
尚学群 (西北工业大学)
申德荣 (东北大学)
孙建伶 (浙江大学)
唐厚君 (North Carolina State University)
邢春晓 (清华大学)
徐俊刚 (中国科学院)
姚 斌 (上海交通大学)
叶新民 (内蒙古大学)
余志文 (华南理工大学)
袁 方 (河北大学)
岳 昆 (云南大学)
张 霞 (东大阿尔派)
张志强 (哈尔滨工程大学)
邹 磊 (北京大学)

简要日程				
日期	时间	程序	地点	报告人/主持人
10月14日	08:00-18:00	注册	一楼主会场外	
	08:30-08:40	研究生学术辅导开幕式	一楼主会场	毛睿
	08:40-09:20	研究生学术辅导1		李建
	09:20-10:00	研究生学术辅导2		陈群
	10:00-10:20	茶歇		
	10:20-11:00	研究生学术辅导3		李荣华
	11:00-11:40	研究生学术辅导4		范举
	12:00-13:30	午餐	前海渔港	
	14:00-17:30	Sigmod China Workshop	一楼主会场	李国良
	18:00-19:30	晚餐	前海渔港	
19:30-22:30	专委会会议	一楼主会场	杜小勇	
10月15日	08:00-18:00	注册	一楼主会场外	毛睿
	08:20-08:50	开幕式	一楼主会场	李国良, 毛睿
	08:50-09:10	合影	一楼主会场外	
	09:10-10:00	大会特邀报告1	一楼主会场	陈国良/彭智勇
	10:00-10:50	大会特邀报告2		余峰
	10:50-11:10	茶歇		
	11:10-11:40	新技术报告1		萧小奎/黄锦辉
	11:40-12:10	新技术报告2		鲍芝峰
	12:10-13:30	午餐	前海渔港	
	14:00-15:40	大数据教学论坛	一楼主会场	李战怀
		软件学报特刊专场	624	崔斌
		分组讨论A	938	范举
		分组讨论B	327	童咏昕
		分组讨论C	328	王宏志
	15:40-16:10	茶歇		
	16:10-17:50	大数据教学论坛	一楼主会场	李战怀
		软件学报特刊专场	624	崔斌, 马帅
		分组讨论A	938	乔少杰
分组讨论B		327	高军	
分组讨论C		328	王朝坤	
18:30-21:30	晚宴	前海渔港		
10月16日	09:00-09:50	大会特邀报告3	一楼主会场	Anhai Doan/李国良
	09:50-10:40	大会特邀报告4		冯玲
	10:40-11:10	茶歇		
	11:10-11:40	新技术报告3		马帅/高宏
	11:40-12:10	新技术报告4		高云君
	12:00-13:40	午餐	前海渔港	
	14:00-14:40	企业报告1	一楼主会场	华为: 何诚/李国良
	14:40-15:20	企业报告2		人大金仓: 冷建全
	15:20-15:40	茶歇		
	15:40-17:00	系统演示	一楼主会场	陈跃国, 伍楷舜
	17:00-17:30	闭幕式&颁奖	一楼主会场	杜小勇
18:00-19:00	晚餐	深大教工餐厅		

特约报告

地点：一楼主会场

主持人：彭智勇

特约报告1：陈国良

10月15日 09:10-10:00

大数据并行与交互式计算

在计算理论中，计算复杂类主要研究P类与NP类两大类问题。在大数据的情况下，为了提高P类问题的求解速度，可采用并行方法，其中一个实例就是NC类并行计算；为了改进NP类问题的求解质量，可采用交互方法，其中一个实例就是IP类交互计算。本报告，首先简要介绍一下预备知识，包括计算模型与计算复杂性，问题的确定性与非确定性求解，P类与NP类基础知识等；然后讨论一下P类问题的并行求解与NP类问题的交互求解；最后在结论中根据目标，研究对策和具体实施方案，概括出大数据计算的总体框架。



陈国良，教授、博士生导师，中国科学院院士，全国首届高等学校教学名师。现任深圳大学计算机与软件学院院长。主要从事并行算法和高性能计算及其应用等研究。先后承担了10多项国家级科研项目。曾获国家科技进步二等奖、国家级教学成果二等奖、2009年度安徽省重大科技成就奖等共20余项。多年来，陈国良教授围绕着并行算法的教学与研究，形成了“算法理论—算法设计—算法实现—算法应用”一套完整的并行算法学科体系，提出了“并行机结构—并行算法—并行编程”一体化的并行计算研究方法，是我国非数值并行算法研究的学科带头人，在国内外学术界和教育界有一定的影响和地位。

特约报告2：余锋

10月15日 10:00-10:50

阿里云数据库研制及其应用

本报告首先阐述云数据库研制的背景、要解决的复杂问题及背后的思考、成为基础设施后带来的挑战，再深入探索工业级别云数据库的架构，核心组件如数据链路、管控系统、高可用、资源隔离、资源调度的实现机制以及中间的设计取舍，最后展现云数据库在各行业的应用情况以及生态系统。



余锋，阿里云研究员，有超过18年的网络和底层系统开发经验，擅长构建大规模集群服务，目前负责阿里云数据库研发和产品体系，在分布式数据库和引擎构建上都有丰富的经验。在VLDB上发表有相关论文，参与多项国家重点科研项目，参与高校教材编写。

特约报告3：Anhai Doan

10月16日 09:00-09:50

Toward a System Building Agenda for Data Integration

Data integration (DI) has long been an important topic in data management, and will become even more so in the age of Big Data and data science. Most DI works so far have focused on developing algorithms. In this talk I argue that we should devote far more effort to building DI systems, in order to truly advance the field. Toward this goal, I begin by drawing on my recent industrial experience to discuss the limitations of current DI systems. Next, I propose an agenda to build a new kind of DI systems to address these limitations. These systems guide users through the DI workflow, step by step. They provide tools to address the “pain points” of the steps, and tools are built on top of the Python data science and big data ecosystem (PyData). I discuss how to foster an ecosystem of such tools within PyData, then use it to build DI systems for collaborative/cloud/crowd/lay user settings. As an example of such systems, I discuss “hands-off” DI systems, which solve the entire DI task using only crowdsourcing. Finally, I discuss ongoing work at Wisconsin, which suggests that these DI systems are highly promising and building them raises many interesting research challenges. If time permits, I will discuss how this DI agenda can be built on to develop a broader R&D agenda for data science.



AnHai Doan is a Professor of Computer Science at the University of Wisconsin-Madison. His interests cover databases, AI, and Web, with a current focus on data integration, data science, Big Data, and data-centric software ecosystems. He received the ACM Doctoral Dissertation Award in 2003, a CAREER Award in 2004, a Sloan Fellowship in 2007, and a Vilas Associate Award in 2016. He co-authored “Principles of Data Integration”, a textbook published by Morgan-Kaufmann in 2012, and is currently serving on the SIGMOD Advisory Board. AnHai has also consulted extensively and been involved in several startups. He was on the Advisory Board of Transformic, a Deep Web startup acquired by Google in 2005, and was Chief Scientist of Kosmix, a social media startup acquired by Walmart in 2011. From 2011 to 2014 he was Chief Scientist of WalmartLabs, a newly formed research and development lab at Walmart, devoted to analyzing and integrating data for e-commerce.

特约报告4：冯玲

10月16日 09:50-10:40

感知预测与疏导青少年心理压力的微博平台

青少年面临巨大压力，超出承受能力会影响身心健康，造成灾难性后果（如伤害自己或他人）。平等、自由、碎片、个性化的微博成为青少年获取信息、沟通表达和释放情绪的重要空间，而微信又为移动青少年用户提供跨媒体、跨平台的即时交流媒介。本报告就如何通过微博，实时感知、预测与疏导青少年心理压力进行探讨。



冯玲，计算机科学与技术系教授，博士生导师，清华大学“百人计划”引进人才，教育部“长江学者”特聘教授，荷兰皇家科学院杰出中青年发明创造奖首位计算机获得者。从事数据管理理论与技术研究30余年，主持承担30余项中国863、自然科学基金、香港研究基金、荷兰科学院基金、澳大利亚研究基金等国家级项目，在国内外重要学术期刊和会议上发表学术论文200余篇。获荷兰皇家Philips、芬兰Nokia专利各一项。提出的情境感知数据管理理论与方法引发未来智能信息社会数据有效管理与服务的一系列相关问题。

新技术报告

地点: 一楼主会场

主持人: 黄锦辉

新技术报告1: 萧小奎

10月15日 11:10-11:40

An Efficient and Accurate Index Structure for SimRank

SimRank is a topology-based similarity measure of nodes in a graph that does not rely on domain-specific information. Scalable SimRank computation has been a subject of extensive research for more than a decade, and yet, none of the existing solutions can efficiently derive SimRank scores on large graphs with provable accuracy guarantees. In particular, the state-of-the-art solution requires up to a few seconds to compute a SimRank score in million-node graphs, and it does not offer any worst-case assurance in terms of the query error. In this talk, I will present an efficient index structure for SimRank that takes near-linear time and space to construct, and that answers any single-pair SimRank query with at most ϵ error in $O(1/\epsilon)$ time. Experiments on real data demonstrate that our solution is orders of magnitude faster than competing methods, while providing higher query accuracy than the latter.



Xiaokui Xiao is an associate professor and Assistant Chair (Strategic Research) at the School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore. His research focuses on data management and data privacy. He received his PhD in Computer Science from the Chinese University of Hong Kong, and was a postdoctoral associate at the Cornell University before joining NTU. He was a recipient of the 2010 Nanyang Assistant Professorship Award from NTU, the 2009 Hong Kong Young Scientist Award in Physical/Mathematical Sciences, and the 2009 ACM-HK Prof. Francis Chin Research Award.

新技术报告2: 鲍芝峰

10月15日 11:40-12:10

A visual exploration of the spatial data

In this talk, I will first introduce the system we build for an interactive and visualized exploration of the location-centered real estate data in Australia for the last ten years, as well as the research problems and techniques behind the system. In particular, we will talk about the query processing to facilitate buyers to find a desired property and for sellers to find the best time to enter the market to sell the house. In the second half of the talk, I will introduce our recent work on exploring the activity trajectory data for interactive trip planning.



Zhifeng Bao is an assistant professor in School of Computer Science & IT, RMIT University, Australia. He received his PhD from the CS Dept at National University of Singapore in 2011. Zhifeng was the recipient of the Best PhD Thesis Award in School of Computing and was the winner of the Singapore IDA (Infocomm Development Authority) gold medal. His research was supported by various researching funding such as Google faculty research award program. He has been committing himself to the task of "how to make data usable", including structured data (e.g. relational data), semi-structured data (e.g. XML), unstructured data (e.g. text), spatial data and graph data (e.g. social network). He focused on building general yet efficient frameworks to support these usability modules, without breaking the traditional storage and indexing scheme for the underlying data. He has served the program committee of top conferences in Database and Information Retrieval such as VLDB, ICDE, SIGIR.

新技术报告3: 马帅

10月16日 11:10-11:40

Towards Big Graph Search: Challenges and Techniques

Graphs have more expressive power and are widely used today, and various applications of social computing trigger the pressing need of a new search paradigm. In this talk, we argue that graph search is the one filling this gap. We first introduce the application of graph search in various scenarios. We then formalize the graph search problem and briefly discuss its challenges. Finally, we introduce several useful query and data techniques towards efficient and effective big graph search.



马帅博士, 北京航空航天大学计算机学院教授、博士生导师; 数据库专业委员会委员, 大数据专家委员会委员; 分别获得了北京大学(2004)和英国爱丁堡大学(2011)的两个博士学位, 英国爱丁堡大学博士后, 曾在美国贝尔实验室总部实习, 在微软亚洲研究院访问。主要从事数据库理论与系统的研究, 2010获数据库领域国际顶级会议VLDB唯一最佳论文; 2012年入选微软亚洲研究院“铸星计划”; 2013年获国家优秀青年科学基金; 2013年获WISE最佳挑战论文奖; 2015年入选微软亚洲研究院“合作研究计划”; 2016年获WSDM CUP挑战赛最终第2名。

新技术报告4: 高云君

10月16日 11:40-12:10

偏好查询结果可用性分析

数据库可用性问题已受到了越来越多专家学者的关注, 尤其是面向查询结果的可用性分析。本报告将汇报报告人所在团队近二年内面向偏好查询结果的可用性问题研究方面的成果(主要包括Why-not & Why问题、Causality & Responsibility问题、Why-few & Why-many问题三方面)。该研究成果可以帮助用户更好地了解查询并得到期望的查询结果, 从而提高数据库的可用性。



高云君, 浙江大学计算机学院教授, 博士生导师, 计算机科学与工程系副主任, 主要研究方向为空间/时空数据库、多源异构(大)数据管理、数据库可用性等。在TODS、VLDBJ、TKDE、TFS、SIGMOD、VLDB、ICDE、SIGIR、EDBT等国内外顶级/重要学术期刊/会议发表论文80余篇, 并获2015年国家优秀青年科学基金、SIGMOD 2015最佳论文提名、ICDE 2015优秀论文、2011年浙江省科学技术一等奖(7/13)。现为CCF数据库专委会委员, CCF高级会员, ACM、IEEE会员, IJSSOE编委, IJDSN专刊特邀编辑, TODS、VLDBJ、TKDE、TMC、TKDD等10余个国内外顶级/重要学术期刊评审员, VLDB、ICDE、SIGSPATIAL GIS、DASFAA等50余个顶级/重要学术会议(共同/出版/宣传/本地)主席/程序委员会委员/评审员。

研究生学术辅导

地点: 一楼主会场

主持人: 毛睿

研究生学术辅导1: 李建

10月14日 08:40-09:20

Some tips for writing high-quality papers.

Paper writing is a crucial skill in academic research. Writing a paper that can be accepted by top conferences or journals is a highly nontrivial task. I have seen many cases where papers with good content but poor writing were rejected (again and again). In this talk, I would like to share with you some tips on how to write good papers, based on others and my own experiences.



Jian Li is currently an assistant professor at Institute for Interdisciplinary Information Sciences (IIIS, previously ITCS), Tsinghua University, headed by Prof. Andrew Yao. He got his BSc degree from Sun Yat-sen (Zhongshan) University, China, MSc degree in computer science from Fudan University (advisors: Hong Zhu, Rudolf Fleischer), China and PhD degree in the University of Maryland, USA (advisors: Amol Desphande and Samir Khuller). His major research interests lie in the broad area of theoretical computer science, in particular algorithm design and analysis, and the algorithmic and theoretical aspects of databases, machine learning and networking. He co-authored several research papers that have been published in major computer science conferences and journals. He received the best paper awards at VLDB 2009 and ESA 2010. He is also a recipient of the "221 Basic Research Plan for Young Faculties" at Tsinghua University and the "new century excellent talents award" by Ministry of Education of China.

研究生学术辅导2: 陈群

10月14日 09:20-10:00

计算机研究之美: 我的经历

学术研究在生理或心理上都给人带来很大的挑战, 通常被初学者认为是一段痛苦的旅程。成功面对这样的挑战关键在于是否能够感受到学术研究之美。我将从我个人的研究经历跟大家分享我对计算机研究之美的理解, 以及如何把痛苦旅程变成一道道曼妙的风景。在讲座中, 我将通过我在数据质量检测与修复、以及大规模并行图处理等方面的一些工作来阐述我的观点。



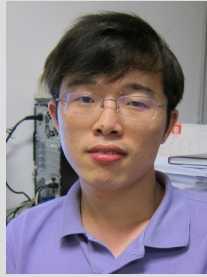
陈群, 清华大学学士, 新加坡国立大学计算机科学博士, 西北工业大学计算机学院教授, 博士生导师, 教育部“新世纪优秀人才计划”和人力资源社会保障部“海外高层次人才回国人员资助计划”的获得者。主要研究领域为数据质量管理和大数据分析等。在国内外知名的学术会议和期刊(包括SIGMOD、ICDE、TKDE 和Information Science等)上发表近60篇文章。主持或作为学术骨干参与多项国家重点科研项目, 包括973、863以及自然科学基金重点项目等。

研究生学术辅导3：李荣华

10月14日 10:20-11:00

图数据库研究的一些体会

在本次报告中，我将和同学们分享我这6年以来在图数据管理领域的一些研究体会。报告将介绍图数据管理应掌握的一些基本算法、如何找到有价值的图数据研究问题、如何分析和解决相关问题。



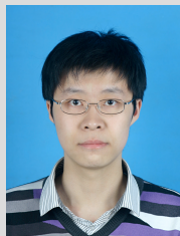
李荣华，博士，深圳大学讲师。李荣华于2013年8月毕业于香港中文大学获得博士学位，同年9月加入深圳大学计算机与软件学院任讲师。李博士的研究兴趣包括大规模图数据管理与挖掘以及其在社交网络分析等领域的应用。迄今为止在上述相关领域发表论文30余篇，其中包括SIGMOD、VLDB、ICDE、KDD、VLDB Journal、TKDE等CCF A类学术会议和期刊论文13篇，第一作者9篇。李博士担任过CCF A类学术会议WWW和CCF B类会议SDM的程序委员会成员，以及长期担任CCF A类期刊ACM TODS和IEEE TKDE的审稿人。

研究生学术辅导4：范举

10月14日 11:00-11:40

众包数据分析：融合机器与人的智能

作为一种新型群体计算模式，众包将复杂计算问题分发给大量的人群进行分布式求解，充分利用人的认知、收集与推理能力参与计算，成功地解决了很多复杂的任务（如实体识别、图像标注等），取得了明显优于传统计算方法的效果，体现出了巨大的应用价值。本次报告侧重介绍众包数据分析——一种融合机器与人智能的大数据分析方法，并具体介绍以下三方面工作：其一是质量控制：由于众包的开放性，其结果往往质量较低，甚至可能出现很大的噪声，因此需要有效的质量控制策略。对此，将介绍一种自适应的众包质量控制策略，综合考虑了众包工人在不同领域上答题精度的差异性，将任务自适应地分配给最擅长的工人进行回答；其二是基于代价的众包查询优化：一个众包查询通常会有多种执行方案，不同方案在执行代价上的差异往往十分巨大。对此，将介绍一种基于代价的众包查询优化方法，综合考虑众包的金钱开销与等待时间；其三是机器众包融合的数据分析：为了保证数据分析的质量与可扩展性，将介绍一种融合机器与众包的方案，在一定的众包预算约束下，设计有效的机器算法，并只在必要的时候引入众包的辅助。同时也会介绍该方法在不同领域，如Web信息集成、知识库构建等的应用案例。在介绍具体工作的同时，也会与研究生听众分享一些自己做研究的感悟与体会。



范举现任中国人民大学副教授，研究领域为数据库系统与数据管理，当前的研究兴趣包括众包数据分析、大数据融合与智能查询等，致力于构建面向多源异构大数据的智能分析系统与工具。范举于2012年在清华大学计算机系获得博士学位，并于2012至2015年在新加坡国立大学任博士后研究员。个人主页：<http://deke.ruc.edu.cn/displaynews.php?id=353>

Sigmod China Workshop

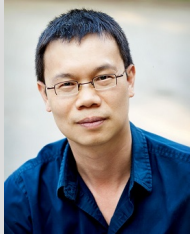
地点: 一楼主会场

主持人: 李国良

报告1: Anhai Doan

10月14日 14:00-14:40

Experience of organizing the Beckman workshop and lessons learned



AnHai Doan is a Professor of Computer Science at the University of Wisconsin-Madison. His interests cover databases, AI, and Web, with a current focus on data integration, data science, Big Data, and data-centric software ecosystems. He received the ACM Doctoral Dissertation Award in 2003, a CAREER Award in 2004, a Sloan Fellowship in 2007, and a Vilas Associate Award in 2016. He co-authored “Principles of Data Integration”, a textbook published by Morgan-Kaufmann in 2012, and is currently serving on the SIGMOD Advisory Board. AnHai has also consulted extensively and been involved in several startups. He was on the Advisory Board of Transformic, a Deep Web startup acquired by Google in 2005, and was Chief Scientist of Kosmix, a social media startup acquired by Walmart in 2011. From 2011 to 2014 he was Chief Scientist of WalmartLabs, a newly formed research and development lab at Walmart, devoted to analyzing and integrating data for e-commerce.

报告2: Anthony K. H. Tung

10月14日 14:40-15:20

Semi-Lazy Learning: Just-in-Time Model Construction Framework and Its Applications

Research on machine learning had started in an era where we have much weaker hardware compared to modern days. In order to ensure that prediction time is acceptable in real-time applications, there are typically two choices in machine learning. First, apply eager learning where a model is pre-constructed before prediction queries arrive and make prediction with the model during the prediction phase. In such a case, the training data is typically abandoned after the model is constructed and thus the model must be sufficiently complex to cater to most of the prediction queries that arrive in a later phase. Second, adopt lazy learning where the data is kept intact and apply simple strategies like k-nearest-neighbors aggregation which have sufficient computational efficiency to produce acceptable throughput. The prediction accuracy of lazy learning can however suffer from their simplistic way of making prediction. In this talk, we will introduce semi-lazy learning, an approach that lie somewhere between eager and lazy learning. Like lazy learning, semi-lazy learning keeps the data intact and construct model only AFTER a prediction query has arrived. The constructed model can be as complex as those that are used in eager learning and such just-in-time model construction is done over a set of selected relevant data. Semi-lazy learning is motivated by the growth in modern hardware which ensures that throughput in prediction is still acceptable even if model is only constructed after prediction queries arrive. We will touch on various techniques that are needed to support semi-lazy learning (www.comp.nus.edu.sg/~atung/gl) and then then introduce applications(eg. ivle.readpeer.com) that we are building on top of the framework.



Anthony K. H. Tung is currently an Associate Professor in the Department of Computer Science, National University of Singapore (NUS). He received both his B.Sc.(2nd Class Honour) and M.Sc. in computer sciences from the National University of Singapore in 1997 and 1998 respectively. In 2001, he received the Ph.D. in computer sciences from Simon Fraser University (SFU). His research group called iData (Intelligence and Data Management Group) has research interest that span across the whole process of converting data into intelligence. Anthony is also the deputy director of NUS SeSaMe research center (<http://sesame.comp.nus.edu.sg/>).

报告3: Xiaokui Xiao**10月14日 15:30-16:10****An Efficient Algorithm for Influence Maximization**

Given a social network G and an integer k , the influence maximization problem asks for k nodes in G that can influence the maximum number of nodes (in expectation) under a stochastic influence propagation model. This problem finds application in viral marketing, and it has been extensively studied over the past decade. Existing methods for the problem, however, either trade the quality of results for computational efficiency, or vice versa. In particular, all methods that ensure $(1 - 1/e - \epsilon)$ -approximate solutions require days to process even a small graph. In this talk, I will present an algorithm for influence maximization that yields both efficiency and accuracy. The algorithm runs in near-linear time and returns $(1 - 1/e - \epsilon)$ -approximate solutions, and it is able to process a billion-edge graph in a few minutes. Experiments on real data demonstrate that our solution is orders of magnitude faster than competing methods, while providing equal or higher accuracy than the latter.



Xiaokui Xiao is an associate professor and Assistant Chair (Strategic Research) at the School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore. His research focuses on data management and data privacy. He received his PhD in Computer Science from the Chinese University of Hong Kong, and was a postdoctoral associate at the Cornell University before joining NTU. He was a recipient of the 2010 Nanyang Assistant Professorship Award from NTU, the 2009 Hong Kong Young Scientist Award in Physical/Mathematical Sciences, and the 2009 ACM-HK Prof. Francis Chin Research Award.

报告4: Jian Li**10月14日 16:10-16:50****Multi-armed Bandits, Online Learning and Sequential Prediction**

Online learning (or sequential prediction) has been a very active research area in machine learning. I will give a very brief introduction to online learning. Then, I will focus on a very popular model in online learning, the stochastic multi-armed bandit model, and present some of our recent results on combinatorial bandits and the pure exploration multi-armed bandit problems.



Jian Li is currently an assistant professor at Institute for Interdisciplinary Information Sciences (IIIS, previously ITCS), Tsinghua University, headed by Prof. Andrew Yao. He got his BSc degree from Sun Yat-sen (Zhongshan) University, China, MSc degree in computer science from Fudan University (advisors: Hong Zhu, Rudolf Fleischer), China and PhD degree in the University of Maryland, USA (advisors: Amol Desphande and Samir Khuller). His major research interests lie in the broad area of theoretical computer science, in particular algorithm design and analysis, and the algorithmic and theoretical aspects of databases, machine learning and networking. He co-authored several research papers that have been published in major computer science conferences and journals. He received the best paper awards at VLDB 2009 and ESA 2010. He is also a recipient of the "221 Basic Research Plan for Young Faculties" at Tsinghua University and the "new century excellent talents award" by Ministry of Education of China.

报告5: C. Mohan

10月14日 16:50-17:30

Hybrid Transaction and Analytics Processing (HTAP): State of the Art

Traditionally, database processing has been broadly classified into two categories: online transaction processing (OLTP) and online analytical processing (OLAP). OLTP systems preceded the emergence of relational database management systems (RDBMSs). OLAP, which was enabled by the arrival of RDBMSs and SQL, and enhancements to them, has gained even more attention in the last decade or so with the emergence of column stores and Big Data technologies like Map/Reduce, Hadoop and Spark. Data generated by OLTP systems are periodically moved in a batched fashion into OLAP systems for analytical processing. In the last few years, increasingly organizations want to be able to base their decisions on the latest set of raw data and the real-time analytics derived from them. This has meant that the capabilities of OLTP and OLAP have had to be combined in a single system with essentially a single copy of the data being used for both purposes. The term HTAP is being used to refer to such Hybrid Transaction and Analytics Processing systems. Currently, there is intense focus on HTAP systems in industry and academia. In this talk, I will discuss the problems, technologies and systems that relate to HTAP. This talk has been given as a keynote during the BIRTE 2016 conference held in conjunction with VLDB 2016.



Dr. C. Mohan has been an IBM researcher for 34 years in the database area, impacting numerous IBM and non-IBM products, the research and academic communities, and standards, especially with his invention of the ARIES family of database locking and recovery algorithms, and the Presumed Abort commit protocol. This IBM (1997), and ACM/IEEE (2002) Fellow has also served as the IBM India Chief Scientist for 3 years (2006-2009). In addition to receiving the ACM SIGMOD Innovation Award (1996), the VLDB 10 Year Best Paper Award (1999) and numerous IBM awards, Mohan was elected to the US and Indian National Academies of Engineering (2009), and was named an IBM Master Inventor (1997). This Distinguished Alumnus of IIT Madras (1977) received his PhD at the University of Texas at Austin (1981). He is an inventor of 47 patents. Recently (2016), he was named a Distinguished Visiting Professor of China's prestigious Tsinghua University. He has served on the advisory board of IEEE Spectrum, and on numerous conference and journal boards. Mohan is a frequent speaker in North America, Europe and India, and has given talks in 40 countries. He is very active on social media and has a huge following. More information could be found in his Wikipedia page at <http://bit.ly/CMwIkP>

大数据教学论坛

地点：一楼主会场

主持人：李战怀

周傲英

10月15日 14:00-14:40

新时代催生新专业：数据科学与工程

大数据的兴起意味着IT发展范式发生了变化。应用驱动创新、开源加速创新、硬件助力创新是这种变化的具体表现。互联网企业IT能力建设的巨大成功，“破除”了传统专业教育带来的“迷信”。对我们而言，现在是实现IT领域弯道超车和跨越式发展的大好时机。有利于我们夯实信息技术基础，形成自己的产品体系，摆脱对垄断企业产品的依赖，实现信息技术产品的“安全可靠、自主可控”目标。为此，需要设置新的数据科学与工程本科专业，培养市场急需的系统架构师和数据科学家。新专业需要在知识结构方面强化计算机、应用数学、信息系统和应用领域的有机融合。为此，课程设置应该加强应用数学训练，裁剪计算机课程，注重信息系统理念的培养。本报告将分享讲者在这方面的思考和实践。



周傲英，华东师范大学副校长、数据科学与工程学院教授。分别于1985和1988年在成都科技大学获得计算机应用学士和硕士学位，1993年在复旦大学计算机系获得博士学位。曾获得国家杰出青年基金和入选长江学者计划特聘教授。目前担任第七届国务院学科评议组成员、中国计算机学会数据库专业委员会副主任、《计算机学报》副主编；曾任ER'2004大会主席、ICDE'2009、ICDE'2012 PC副主席、VLDB'2014 PC共同主席。研究兴趣主要包括Web数据管理、数据密集型计算、内存集群计算、分布事务处理，大数据基准测试和性能优化。

王建民

10月15日 14:40-15:20

大数据硕士培养实践

清华大学在研究生层次以第二学位模式已经开展多年教育活动，本报告将分享清华大学在大数据硕士培养方面的实践经验。



王建民，男，教授，清华大学软件学院副院长、党委书记，全国信息安全标准化技术委员会大数据安全标准特别工作组组长，国家重大科技专项“核高基”基础软件方向实施专家组成员，国家“十二五”863计划先进制造技术领域制造服务技术主题专家组成员，国家“十二五”863计划先进制造技术领域“面向制造业的核心软件开发”重大项目总体专家组组长，国家卫计委人口健康信息化专家咨询委员会委员，ISO/IEC JTC1 大数据研究组（SGBD）成员，IEEE 过程挖掘任务组(IEEE Task Force on Process Mining)成员，中国计算机学会数据库专委会委员，Petri网专委会委员，大数据专委会委员，《计算机集成制造系统》杂志(EI源刊)副主编。获国务院颁发的政府特殊津贴，国家科技进步二等奖（排名第一），入选国家“万人计划”科技创新领军人才，中国电子学会科技进步一等奖（排名第一），入选国家科技部中青年科技领军人才，教育部科技进步一等奖（排名第一），教育部科技进步二等奖（排名第一），国家科技进步二等奖(排名第四)，北京市师德先进个人，清华大学研究生良师益友名人堂，清华大学研究生良师益友，清华大学先进工作者。

地点：一楼主会场

主持人：李战怀

张祖平

10月15日 15:20-16:00

数据科学与大数据技术专业申请与学科同步建设

中南大学于2015年正式创办计算机科学与技术专业大数据方向，2015年7月正式申报教育部“数据科学与大数据技术专业”，并成为获得教育部第一批批准设立“数据科学与大数据技术专业”的国内三所高校之一。作为全程参与“数据科学与大数据技术专业”申报的主要负责人员之一，讲者将与大家分享这个过程的经验与感受。专业要发展，离不开学科与平台的支撑，随着“湖南省医学大数据协同创新中心”平台的建立及医学大数据、交通大数据、银行大数据等项目的实施，计算机与医学、交通、金融等专业领域的交叉融合所产生的模型、方法及技术等，将为大数据专业未来建设发展提供强有力的支撑，实现大数据专业人才培养与改善人才培养环境的双向互补，讲者将与大家分享这方面的一点想法。



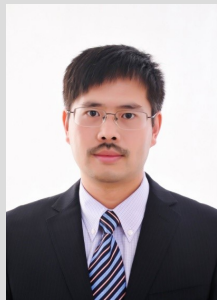
张祖平，湖南师范大学基础数学本科，吉林大学基础数学硕士，中南大学计算机应用技术博士，基础医学博士后（在湘雅医学院肿瘤研究所从事基因数据分析与处理），加拿大西安大略大学(The University of Western Ontario)国家公派访问学者。中南大学信息与工程学院教授，博士生导师，计算机科学与技术系主任，大数据与知识工程研究所副所长。研究兴趣主要包括大数据与网脑，信息度量与信息融合、软件工程与信息系统、参数计算与生物计算。

林子雨

10月15日 16:20-17:00

以平台化思维构建中国高校大数据课程公共服务体系

为了加快中国高校大数据课程体系建设，促进中国高校大数据教学水平不断提升，迫切需要以平台化思维构建中国高校大数据课程公共服务体系，从而获得中国高校大数据教学事业的长足发展。在本报告中，讲者将着重分析建立高校大数据课程公共服务体系的重要性，指出了以平台化思维构建中国高校大数据课程公共服务体系的基本要求，并介绍一个典型案例——“中国高校大数据课程公共服务平台”。



林子雨，男，1978年出生，博士（毕业于北京大学），现为厦门大学计算机科学系助理教授，海峡云计算与大数据应用研究中心副主任，厦门大学数据库实验室负责人，厦门大学云计算与大数据研究中心主要建设者和骨干成员。曾任厦门大学信息科学与技术学院院长助理、晋江市发展和改革局副局长。中国高校首个“数字教师”提出者和建设者，2009年至今，“数字教师”大平台累计向网络免费发布超过100万字高价值的研究和教学资料，累计网络访问量超过100万次。编著出版了国内高校第一本系统介绍大数据知识的专业教材《大数据技术原理与应用》，并成为京东、当当网等网店畅销书籍。建设了国内高校首个大数据课程公共服务平台，为教师教学和学生学习大数据课程提供全方位、一站式服务，年访问量超过50万次；建设了国内高校首个大数据课程教师培训交流基地，为全国高校培养大数据课程师资力量，2016年8月到9月，顺利完成3期培训，总计有来自全国60多所高校的100余位教师参加了培训交流。主要研究方向为数据库、数据仓库、大数据，并以第一作者身份在《软件学报》《计算机学报》和《计算机研究与发展》等国家重点期刊以及国际学术会议上发表多篇学术论文。作为项目负责人主持的科研项目包括1项国家自然科学基金项目(No.61303004)、1项福建省自然科学基金项目(No.2013J05099)和1项中央高校基本科研业务费项目(No.2011121049)，同时，作为课题负责人完成了国家发改委城市信息化重大课题、国家物联网重大应用示范工程区域试点泉州市工作方案、2015泉州市互联网经济调研等课题。

10月15日 17:00-18:00

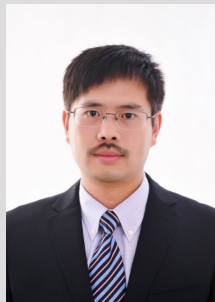
嘉宾观众互动问答



李翠平,女,教授,中国人民大学信息学院副院长、计算机系主任,中国计算机学会杰出会员,大数据专家委员会委员,数据库专委会委员。数据工程与知识工程教育部重点实验室兼职研究员,国际会议VLDB(2012-2015)、ICDE2012等程序委员,TKDE、VLDB Journal等期刊匿名评审专家。2003年毕业于中科院计算所获得博士学位,2008年在美国UIUC大学计算机系访问交流。目前研究方向为数据仓库、数据挖掘、社会网络分析和社会媒体推荐等。主持和参与国家自然科学基金、973、863等10多项国家级和省部级项目,先后在SIGKDD、SIGMOD、VLDB、EDBT等重要国际会议和TKDE、KAIS、JCST、软件学报等国内外期刊发表论文100多篇,出版著作2部,申请国家发明专利4项,获得软件著作权登记2项。2007年入选北京市组织部优秀人才培养资助计划。2009年入选教育部新世纪优秀人才支持计划。2011年获得中国计算机学会“青年科学家奖”。



张祖平,湖南师范大学基础数学本科,吉林大学基础数学硕士,中南大学计算机应用技术博士,基础医学博士后(在湘雅医学院肿瘤研究所从事基因数据分析与处理),加拿大西安大略大学(The University of Western Ontario)国家公派访问学者。中南大学信息与工程学院教授,博士生导师,计算机科学与技术系主任,大数据与知识工程研究所副所长。研究兴趣主要包括大数据与网脑,信息度量与信息融合、软件工程与信息系统、参数计算与生物计算。



林子雨,男,1978年出生,博士(毕业于北京大学),现为厦门大学计算机科学系助理教授,海峡云计算与大数据应用研究中心副主任,厦门大学数据库实验室负责人,厦门大学云计算与大数据研究中心主要建设者和骨干成员。曾任厦门大学信息科学与技术学院院长助理、晋江市发展和改革局副局长。中国高校首个“数字教师”提出者和建设者,2009年至今,“数字教师”大平台累计向网络免费发布超过100万字高价值的研究和教学资料,累计网络访问量超过100万次。编著出版了国内高校第一本系统介绍大数据知识的专业教材《大数据技术原理与应用》,并成为京东、当当网等网店畅销书籍。建设了国内高校首个大数据课程公共服务平台,为教师教学和学生学习大数据课程提供全方位、一站式服务,年访问量超过50万次;建设了国内高校首个大数据课程教师培训交流基地,为全国高校培养大数据课程师资力量,2016年8月到9月,顺利完成3期培训,总计有来自全国60多所高校的100余位教师参加了培训交流。主要研究方向为数据库、数据仓库、大数据,并以第一作者身份在《软件学报》《计算机学报》和《计算机研究与发展》等国家重点期刊以及国际学术会议上发表多篇学术论文。作为项目负责人主持的科研项目包括1项国家自然科学基金青年基金项目(No.61303004)、1项福建省自然科学基金青年基金项目(No.2013J05099)和1项中央高校基本科研业务费项目(No.2011121049),同时,作为课题负责人完成了国家发改委城市信息化重大课题、国家物联网重大应用示范工程区域试点泉州市工作方案、2015泉州市互联网经济调研等课题。

企业报告

地点: 一楼主会场

主持人: 李国良

何诚

10月16日 14:00-14:40

机器学习和人工智能使能智慧网络运维管理

随着Internet和移动宽带的高速发展,以及丰富多样的新业务出现,网络流量呈现出爆炸性增长的趋势,这对网络领域深度分析带来无限可能的同时也带来了极大的挑战。通过引入大数据、机器学习和人工智能相关技术,并结合网络领域知识和业务进行优化,可以极大提升网络运维管理甚至控制效率和准确度。



何诚,2006年加入华为公司中央研究院,现任诺亚方舟实验室研究员,主任工程师,人工智能在智能网络运维管理领域研究负责人。研究领域包括大规模网络流量测量与建模,大数据实时流分析,智能网络运维管理等。

冷建全

10月16日 14:40-15:20

金仓数据库性能优化方法与实践

性能优化是软件技术发展过程中的永恒话题,理论界和工业界都在此领域投入了大量的精力。人大金仓起源于人民大学信息学院,是国内最早从事数据库系统产品研发的企业。在人大金仓30余年的数据库实现技术与开发过程中,人大金仓总结出了一套自己的数据库性能优化方法,并在自身的产品研发过程中得到了大量实践检验。本次演讲,我们将分享我们在此领域的思考和实践经验,希望与各位理论界的专家共同探讨。



冷建全,毕业于中国人民大学信息学院,获工学硕士学位。毕业后加入人大金仓,现为人大金仓公司数据库业务部总监,主要负责金仓数据库的架构设计和研发。对数据库实现技术有深入研究,拥有多年的研究和开发经验。

获奖论文

第33届中国数据库学术会议最佳论文

云存储中支持失效文件快速查询的批量审计方法

Huifeng Wang, Zhanhuai Li, Xiao Zhang, Jian Sun and Xiaoan Zhao

第33届中国数据库学术会议萨师焯优秀论文

基于Multi-agent的分布式文本聚类模型

Shaojie Qiao and Nan Han

HybriG: 一种高效处理大量重边的属性图存储架构

黄权隆, 黄艳香, 邵荃侠, 孟嘉, 任鑫琦, 崔斌 and 冯是聪

基于相异支撑点选取算法的度量空间离群检测

Honglong Xu, Xingliang Li, Jing Shen and Gang Liu

基于众包的知识库索引对齐算法

Bingwen Shen and Jianhua Feng

基于蚁群算法的Map-Reduce环境下的多连接查询优化方法

Yue Mingliang, Gao Hong and Wang Hongzhi

一种新的用于跨领域推荐的迁移学习模型

Jun Wang, Shijun Li, Sha Yang and Wei Yu

路网空间下基于马尔可夫决策过程的异常车辆轨迹检测算法

Jiangyun Mao, Hao Wu and Weiwei Sun

对象代理数据库的双向指针存储优化方法

胡聪睿 and 彭智勇

基于DTW的时序数据相似连接

Ningnan Zhou, Xiao Zhang and Shan Wang

软件学报特刊专场

10月15日

地点: 624

主持人: 崔斌, 马帅

14:00-15:40

16:10-17:50

融合主题模型和协同过滤的多样化手机应用推荐

黄璐, 刘红岩, 何军, 杜小勇

基于深度稀疏自动编码器的社区发现算法

尚敬文, 王朝坤, 辛欣, 应翔

基于社交关系的微博主题情感挖掘

黄发良, 于戈, 李超雄

社交网络环境下基于信任的推荐算法

陈婷, 朱青, 周梦溪, 王珊

复杂网络大数据中重叠社区检测算法

乔少杰, 韩楠, 张凯峰, 邹磊, 王宏志, Louis Albert

一种面向HDFS的多层索引技术

何龙, 陈晋川, 杜小勇

动态信息网络中基于角色的结构演化与预测

李川, 冯冰清, 李艳梅, 胡绍林

空间众包环境下的三类对象在线任务分配

宋天舒, 童咏昕, 王立斌, 许可

八叉树立体网格Hilbert曲线索引算法

曹雪峰

An Automatic Web Data Extraction Approach based on Path Sub-index Tree

Yan Wen, Bo Zhang, Qingtian Zeng and Ming Chen

一种基于众核架构Phi协处理器的内存OLAP外键连接算法

张宇, 张延松, 陈红, 王珊

集群数据库系统的日志复制和故障恢复

王嘉豪, 蔡鹏, 钱卫宁, 周傲英

分布式数据流上的高性能发策略

房俊华, 王晓桐, 张蓉, 周傲英

透析计算: 面向OLGP的InfoNetCube高效物化

刘光明, 李川, 唐常杰

基于语义约束LDA的商品特征和情感词提取

彭云, 万常选, 江腾蛟, 刘德喜, 刘喜平, 廖国琼

一种云环境中数据流的高效多目标调度方法

沈尧, 秦小麟, 鲍芝峰

面向多源大数据云端处理的成本最小化方法

肖文华, 包卫东, 朱晓敏, 邵屹, 杨陈超, WuJianhong

基于最小费用大流的规模资源调度方法

陈晓旭, 吴恒, 悦文, 陆志刚, 张博

MapReduce 大数据处理平台与算法进展

宋杰, 孙宗哲, 毛克明, 鲍玉斌, 于戈

基于小波密度估计的RFID离群点检测

刘耀宗, 刘云恒, 于继明

分组A

10月15日

地点: 938

14:00-15:40

主持人: 范举

16:10-17:50

主持人: 乔少杰

一种基于差分隐私的频繁序列挖掘算法
Yanhui Li, Hao Liu, Ye Yuan and Guoren Wang

基于多级空间值的真值发现算法
刘心乔, 申德荣, 聂铁铮, 寇月 and 于戈

HybriG: 一种高效处理大量重边的属性图存储架构
黄权隆, 黄艳香, 邵莹侠, 孟嘉, 任鑫琦, 崔斌 and 冯是聪

KSRG: 一种高效的多关键词匹配最优路径查询算法
Jin Pengfei, Niu Baoning and Zhang Xingzhong

多元化top-k shapelets分类方法
孙其法, 闫秋艳 and 闫欣鸣

Helius: A Lightweight Big Data Processing System
Mengsu Ding and Shimin Chen

基于相异支撑点选取算法的度量空间离群检测
Honglong Xu, Xingliang Li, Jing Shen and Gang Liu

一种基于领域驱动的高效用Co-location模式挖掘方法
Wanguo Jiang, Lizhen Wang and Yuan Fang

基于众包的知识库索引对齐算法
Bingwen Shen and Jianhua Feng

一种基于Storm的大数据环境下交通流在线预测算法
Jimlove Jim, Dianchao Zhang and 王汝传

基于蚁群算法的Map-Reduce环境下的多连接查询优化方法
Yue Mingliang, Gao Hong and Wang Hongzhi

面向评价数据中用户偏好发现的隐变量模型构建与推理
Yan Gao, Kun Yue, Hao Wu, Xiaodong Fu and Weiyi Liu

路网空间下基于马尔可夫决策过程的异常车辆轨迹检测算法
Jiangyun Mao, Hao Wu and Weiwei Sun

LSH indexing based on optimized linear order of keys
Wenfei Jiao, Xiaokang Feng, Xiaoyu Zhang and Jiangtao Cui

对象代理数据库的双向指针存储优化方法
胡聪睿 and 彭智勇

基于模糊粗糙集的属性约简增量方法研究
Li Liu and Suyun Zhao

一种新的用于跨领域推荐的迁移学习模型
Jun Wang, Shijun Li, Sha Yang and Wei Yu

基于Multi-agent的分布式文本聚类模型
Shaojie Qiao and Nan Han

基于DTW的时序数据相似连接
Ningnan Zhou, Xiao Zhang and Shan Wang

基于查询概率的位置隐私保护方法
Dapeng Zhao

Research of Distributed Clustering Storage Technology For Massive Remote Sensing Data
Ji Yan, Lu Kewen and Zhang Yinghui

云存储中支持失效文件快速查询的批量审计方法
Huifeng Wang, Zhanhuai Li, Xiao Zhang, Jian Sun and Xiaoan Zhao

面向基因数据分类的核主成分分析RoF算法研究
陆慧娟, 刘亚卿, 孟亚琼 and 郑文斌

The Worm Detection Approach Based on Cloud Computing
Hanxun Zhou

一种基于中医方剂数据库的Top-Rank-k频繁模式挖掘算法
谭龙 and 秦琦冰

基于时间和频次加权张量分解的个性化标签推荐系统
Keqiang Wang, Yuanyuan Jin, Hongwei Peng and Xiaoling Wang

Unified POI Recommendation Method Research Based on Spark
Weizhong Xue, Kankan Zhao, Cuiping Li, Hong Chen, Hui Sun and Suyun Zhao

道路网络上基于时空相似性的连续查询隐私保护算法
潘晓, 谌伟璋, 孙一格

基于张量分解与重建的多因素影响程度分析方法
王卿云, 李红燕, 洪申达

Exploring the Industrial Internet structure: a case study on supply chain networks

Jing Shen, Hao Liao, Yifan Mao, Zecheng Huang and Mingyang Zhou

A Quality Ranking Algorithm for Online Systems based on Fluid Dynamics Theory

Zecheng Huang, Hao Liao, Yifan Mao, Jing Shen and Kezhong Lu

基于Spark的Top-k对比序列模式挖掘
张鹏, 段磊, 秦攀, 左劼, 唐常杰

多准则赋权排序与C-SVM相结合的特征选择算法
孙勤, 蒋艳凰, 胡维, 张毅

基于Storm的决策树分类算法并行化研究

Jimlove Jim, Yongpan Zhang, Haiyu Tan and 王汝传

支持得分矩阵的音乐检索技术

王培培, 王斌

基于历史查询的差分隐私流数据自适应发布算法

Yingjie Wu, Liqun Zhang, Jian Kang and Yilei Wang

一种基于MPI的大规模栅格影像并行瓦片化算法

刘世永, 吴秋云, 陈萃, 熊伟, 吴焯 and 李军

基于概率图模型的用户评价数据分析和动态行为建模

王飞, 岳昆, 孙正宝, 武浩, 冯辉

基于弱水印的轨迹数据完整性检验方法

文春雷, 汪传建 and 王伟强

一种基于SPARK的多路空间连接查询处理算法

乔百友, 朱俊海, 郑宇杰, 申木川, 王国仁

基于联系度的主题关注网络社区发现方法研究

陈晓, 郭景峰, 范超智

实时轨迹隐私保护方法的研究

Chunhe Liao, Zhigang Zhang, Xiuxia Tian, Cheqing Jin and Xiaoyi Duan

基于Storm的流数据去重计数HyperLogLog算法改进研究

Jimlove Jim, Dianchao Zhang and Ruchuan Wang

Influence maximization in social network under budget

Liu Yuanying, Jingfeng Guo and Wei Lidong

基于对称索引的增量式重叠区间连接算法

Zenglong Wang and Jizhou Luo

面向海量交通数据的HBase时空索引研究

Jun Fang, Dong Li and Yanbo Han

云环境中基于代理重加密的多用户全同态加密方案

李陶深, 刘青 and 黄如维

Probabilistic Bichromatic Reverse-kNN query on Road Network

Xu Wei, Li Wengen, Zhang Yichao and Guan Jihong

GMVI: An Online Gyro-shaped Missing Values Imputation for Sensor Data

Jianguo Zheng, Yu Gu, Fangfang Li and Ge Yu

基于时序画像图的疾病风险预测方法

Shuai Zhang and Lizhen Cui

Adaptive Strategy based Dynamic Acquisition Method of User Interest

Xu Wu

基于z值的分布式密度峰值聚类算法

卢晶, 段勇, 刘海波

基于Spark的方体计算与存储

史少晨, 赵衍衍, 李翠平, 陈红, 孙辉 and 邓索

基于复杂网络的工业供应链分析

沈婧, 廖好, 毛一帆, 黄泽成, 周明洋

多样性感知的时空文本信息的kNN查询处理方法

李晨, 申德荣, 寇月, 聂铁铮 and 于戈

面向流数据的决策树分类算法并行化研究

季一木, 张永潘, 郎贤波, 张殿超, 王汝传

基于Storm的流数据管理客户端交互系统设计与实现

陆莉莉, 张殿超, 谈杰, 季一木

GAPI: 一种GPU加速的移动对象并行索引方法

Qingshou Che, Chuanwen Li, Yi Zhang and Qingxu Deng

基于KD-Tree的隐私空间分割方法

Kaizhong Jin and Xiaojian Zhang

面向可视化的全局自适应等距映射算法

Xiyang Wang, Chunling Cheng and Xingguo Chen

Big Data Similarity Join Processing Based on Prefix-Suffix Filtering

Shizhuo Deng, Junchang Xin, Tiezheng Nie and Guoren Wang

Crypt-JDBC model: Optimization of Onion Encryption Algorithm

He Chen, Xiuxia Tian and Cheqing Jin

基于Spark平台的短文本特征扩展与分类研究

王雯, 赵衍衍, 李翠平, 陈红 and 孙辉

面向大数据的流分类挖掘算法及其概念漂移应用研究

Lili Lu, Yongpan Zhang, Haiyu Tan and Jimlove Jim

支持近似图查询的Why-Not问题解释方法

贺丹, 宗传玉, 王斌, 李金旭 and 杨晓春

基于贝叶斯网的知识图谱链接预测

Lu Han, Zidu Yin, Yujie Wang, Kuang Hu and Yue Kun

基于变结构动态贝叶斯网络的交通流量预测

Yongheng Wang, Hui Gao and Xuanling Chen

数据流中 ρ -支配轮廓查询算法

王之琼, 霸建民, 黄达 and 信俊昌

基于动作日志时间最优的利润最大化

Shengnan Xie and Yong Liu

大气再分析产品应用平台通用框架及组件设计

朱祥茹, 刘柏年, 张卫民

改进RSF主动轮廓模型的医学图像分割方法

郑彦, 元昌安, 覃晓 and 周凯

NoSQL上高效Top-k范围聚合查询

朱锋, 刘杰, 许利杰, 伍海江, 叶丹, 魏峻, 黄涛

一种高效的降低语义轨迹基础频繁模式稀疏度的方法

Xia Wu, Zhiyong Peng, Yuwei Peng and Yuanyuan Zhu

Efficient Mining of a new representation of high-utility sequences

Shiming Guo, Hong Gao and Jianzhong Li

基于Storm的DSMS交互系统客户端的设计与实现

Jimlove Jim, Dianchao Zhang, Jie Tan and Ruchuan Wang

基于信息网模型的分布并行多连接查询优化

徐晶, 刘梦赤

基于顺序检测的阈值为k的近似串匹配

牛之贤, 张润梁, 牛保宁

公共交通网络下关键词覆盖的最优路径搜索

曹伍, 戴戈南, 徐蓓, 刘玉葆

基于H指数的含时在线网络排序算法研究

毛一帆, 廖好, 黄泽成, 沈婧, 陆克中

基于分层抽样的数据流近似查询算法

倪赛龙, 王永利, 赵忠文, 董振江

IM2: 一种改进的MIN/MAX窗口函数优化技术

宋光旋, 赵大鹏, 王晓玲

A Two-phases Frequent Itemsets Mining Algorithm on SPARK

Yande Zhao, Wei Zhang, Hong Gao and Jianzhong Li

密度聚类划分时间段的动态热度路网构建

周宇鹏, 牛保宁

一种分类数据聚算法及其高效并行实现

丁祥武, 谭佳, 王梅

系统演示

10月16日 15:40-17:00

地点：一楼主会场

主持人：陈跃国，伍楷舜

Libonev: 一个面向可扩展数据库管理系统的通信库
肖冰, 郭进伟, 钱卫宁, 周傲英

一个面向中文的看图造句系统
蓝玮毓, 王晓旭, 李锡荣

ApMed: 面向大规模医学图像的自动化预处理系统
李文博, 潘海为, 谢晓芹, 张志强, 冯晓宁

基于Storm的在线产品评论数据的实时采集系统
郝志峰, 骆魁永, 蔡瑞初

Hadoop大数据平台的统一数据资源管理系统
胡俊峰, 王鹏飞, 谢志鹏, 何震瀛, 陈诚

日志数据的无丢失暂存与快速装载
覃雄派, 金国栋, 刘阳, 丛一鸣, 杜小勇

TCARS: 基于主题模型的测试用例自动推荐系统
母红芬, 李征, 焦永强, 程新宇

P2PRsy: P2P借贷平台贷款风险评估与推荐系统
王耀辉, 杨梦迪, 王睿涵, 张一舟, 熊贇, 米晋宏

KECVS: 一个面向专业文献知识实体的类型标注及可视化系统
温雯, 伍思杰, 蔡瑞初, 郝志峰, 王丽娟

DRSQL: 基于分布式数据库中间件的Document-Relation关联查询系统
薛皓, 黄晓虎, 王梅

HeteVis: 基于交互式中心度抽样算法的异质网络可视化分析系统
赵健霏, 陈晓纪, 肖丁, 石川

一种基于半监督学习方法的问答对广告判定系统及演示
陈修司, 李笑宇, 高军

NEUERS: 可扩展的实体识别系统
孙琛琛, 汪潜, 亓建顺, 王习特, 申德荣, 于戈

医学图像Top-k查询系统
王保加, 潘海为, 谢晓芹, 张志强, 冯晓宁, 韩启龙

支持多核平台的Esper数据流处理系统
刘杰, 王洪亚, 陆可镜

支持多模态医学影像数据融合的建模与并行加载系统
翟霄, 潘海为, 谢晓芹, 张志强, 韩启龙

中文领域知识图谱构建系统
樊艳, 汪诚愚, 何晓丰, 周国民

思眠: 智能睡眠监测与调节系统
陈东, 张禹, 任春霞, 李金宝