

大会程序概览

2013年8月16日（星期五）

9:00-22:00	报到注册（八一宾馆、鑫城宾馆）
12:00-13:30	中餐（八一宾馆二楼）
13:30-15:10	研究生学术辅导 1+2 （八一宾馆多功能报告厅） 主持人：韩启龙
15:10-15:30	茶歇
15:30-17:10	研究生学术辅导 3+4 （八一宾馆多功能报告厅） 主持人：韩启龙
18:00-19:00	晚餐（八一宾馆二楼）
20:30-22:30	数据库专委会会议 （八一宾馆多功能报告厅） 主持人：周立柱

2013年8月17日（星期六）

8:00-12:00	报到注册（启航活动中心）	
8:30-9:00	开幕式（启航剧场） 主持人：印桂生	
9:00-9:20	集体合影	
9:20-10:10	特邀报告 1（李建中） 主持人：周立柱	
10:10-10:30	茶歇	
10:30-11:20	特邀报告 2（樊文飞） 主持人：唐常杰	
11:20-12:10	特邀报告 3（吴 枫） 主持人：孟小峰	
12:20-13:30	午餐（八一宾馆二楼）	
14:00-14:45	新进展报告 1（林学民）	启航剧场
14:45-15:30	新进展报告 2（杜小勇）	主持人：印桂生
15:30-15:50	茶歇	
15:50-16:35	新进展报告 3（周傲英）	启航剧场
16:35-17:20	新进展报告 4（李战怀）	主持人：高宏

18:00-20:00 大会晚宴（红事会）

2013年8月18日（星期日）

8:30-10:00	分组报告-1（启航活动中心）			
	长江厅	黄河厅	黄海厅	西海厅
	分组 A1	分组 B1	分组 C1	分组 D1
10:00-10:20	茶歇			
10:20-11:35	分组报告-2（启航活动中心）			
	长江厅	黄河厅	黄海厅	西海厅
	分组 A2	分组 B2	分组 C2	分组 D2
12:00-13:30	午餐（八一宾馆）			
14:00-14:30	企业报告 1（李丽虹）	启航剧场 主持人：张志强		
14:30-15:00	企业报告 2（冯是聪）			
15:00-15:30	企业报告 3（段刚）			
15:30-15:50	茶歇			
15:50-16:30	企业报告 4（李航）	启航剧场 主持人：于旭		
16:30-17:00	企业报告 5（黎文宪）			
17:00-17:30	闭幕式（启航剧场） 主持人：周立柱			
18:00-19:00	晚餐（八一宾馆二楼）			

大会开幕式

地点：启航剧场

时间：2013年8月17日 8:30-9:00

主持人：印桂生

特邀报告

地点：启航剧场

特邀报告 1	大数据计算：基本概念、研究问题与部分解
时间	2013年8月17日 9:20-10:10
主持人	周立柱
	李建中 哈尔滨工业大学
<p>个人简介：</p> <p>李建中，哈尔滨工业大学教授、博导，1995 年获得国家杰出青年基金，现任 973 项目首席科学家，中国计算机学会常务理事、传感器网络专业委员会主任、大数据专家委员会副主任，曾任《IEEE Transactions on Knowledge and Data Engineering》等国际著名期刊副主编，《计算机学报》、《软件学报》等国内一流刊物编委。</p> <p>他长期从事海量数据计算、传感器网络等方面的研究工作，在国内外著名学术会议和刊物发表论文 200 余篇，其中在 SIGMOD、VLDB、INFOCOM、ICDE、Algorithmica、VLDB Journal、IEEE Transactions on Knowledge and Data Engineering、IEEE Transactions on Parallel and Distributed Systems 等国际计算机顶级学术会议和学术刊物发表论文近 70 余篇，被他引 11500 余次，单篇论文最高他引 1500 余次，三篇论文获得国际会议最佳论文奖，其中一篇获国际计算机顶级会议 VLDB 最佳论文奖。</p> <p>应用基础研究成果，他于 1993 年研制出我国第一个微型计算机集群并行计算系统（云集算系统雏形），1998 年研制出我国第一个机群并行数据库系统，在很多领域得到应用，取得了很大经济和社会效益。研究成果分别获得国家科技进步二等奖 1 项、省部级自然科学一等奖 1 项、省部级科技进步一等奖 2 项，其他奖励 10 余项。</p> <p>报告摘要：</p> <p>该报告首先从计算科学的角度讨论大数据计算的基本概念，包括大数据、大数据计算、大数据计算问题空间等概念。然后，该报告讨论大数据计算的研究问题，包括两个方面的基础理论问题、六个方面的关键技术问题以及三个困惑的问题。在讨论研究问题的同时，该报告还介绍报告人的部分相关工作和相关研究结果。</p>	
特邀报告 2	Querying Big Data
时间	2013年8月17日 10:30-11:20

主持人	唐常杰
	Wenfei Fan University of Edinburgh, UK
<p>个人简介:</p> <p>樊文飞, 英国爱丁堡大学信息学院主任教授, 英国爱丁堡/苏格兰皇家科学院院士, 美国计算机协会会士 (ACM Fellow), 中国“千人计划”特聘专家, 长江学者。他毕业于北京大学 (本科, 硕士) 和美国宾夕法尼亚大学 (博士)。曾获得英国“罗杰·尼达姆奖” (2008年)、Alberto O. Mendelzon 时间检验奖 (ACM PODS 十年最佳论文奖, 2010年)、VLDB (2010年) 和 ICDE (2007) 最佳论文奖、Elsevier 网络科学刊物年度最佳论文和最杰出作者奖 (2002年) 和美国 CAREER Award (2001年) 等多项国际奖项。目前主要研究领域为数据库理论与系统, 包括大数据、数据质量、数据集成、分布式查询处理、查询语言、推荐系统、社会网络查询与分析, 与 Web 服务等方面。</p>	
<p>报告摘要:</p> <p>Big data poses new challenges to query answering, from computational complexity theory to query evaluation techniques. Several questions arise. What query classes can be considered tractable in the context of big data? How can we make query answering feasible on big data? Is MapReduce the best way for answering queries on big data? What should we do about the quality of the data, the other side of big data? This talk aims to provide an overview of recent advances in tackling these questions.</p>	

特邀报告 3	Mediabase
时间	2013年8月17日 11:20-12:10
主持人	孟小峰
	吴枫 微软亚洲研究院资深研究员
<p>个人简介:</p> <p>吴枫于 1992 年在西安电子科技大学获学士学位, 并分别于 1996 年和 1999 年在哈尔滨工业大学获得硕士和博士学位。随后他加入了微软亚洲研究院, 担任助理研究员, 2001 年升任研究员, 现为资深研究员/研究经理。其研究兴趣包括图像与视频压缩, 多媒体传输, 多媒体分析与合成, 发表了超过 200 篇高水平学术论文 (包括 50 余篇国际期刊论文), 申报了 67 项美国专利 (其中 25 项获得授权), 13 项技术被国际视频编码标准所采纳, 获得了 2009 年 IEEE T-CSVT, 2008 年 PCM 以及 2007 年 SPIE VCIP 的最佳论文奖。他是 IEEE Fellow, 担任 IEEE T-CSVT、IEEE T-MM 以及其他国际期刊编委, 担任 2011 年的 MMSP、2010 年 VCIP 以及 2009 年 PCM 技术委员会主席。</p>	
<p>报告摘要:</p> <p>Image and video have become the majority of big data in quantity. In this talk, I will first introduce our two new researches on large-scale images. The first one proposes a novel album</p>	

coding to significantly reduce the storage size in the cloud. The key idea is to exploit the correlation of images and organize them as a pseudo sequence for compression. The second one proposes the cloud-based image coding, which is different from current image coding even on the ground. It no longer compresses images pixel by pixel and instead tried to describe images and reconstruct them from a large-scale image database via the descriptions. Motivated by these two researches, it is time to study how to organize and store large-scale images, including metadata and image data. The targets include not only efficient storage but also easy for various applications. We called this work as Mediabase.

新进展报告

地点：启航剧场

主持人：印桂生、高宏

新进展报告 1	Recent Research Progress of UNSW DBG in Network Data based Search
时间	2013 年 8 月 17 日 14:00-14:45
	林学民 新南威尔士大学
<p>个人简介：</p> <p>林学民，新南威尔士大学计算机科学及工程学院教授、数据库研究实验室主任。林学民教授长期从事数据库理论、算法与技术研究，主要工作领域是时空数据和流数据的查询、图和文本的匹配查询、不确定数据的概化查询及图数据可视化等。林学民教授目前是 ACM Transactions on Database Systems 和 IEEE Transactions on Knowledge and Data Engineering 的编委，并多次应邀担任 SIGMOD, VLDB, ICDE, KDD 等权威会议的程序委员。近年来在数据库等相关领域的顶级国际学术会议 (SIGMOD, VLDB, KDD, ICDE, SIGIR) 与国际顶级期刊 (ACM TODS, VLDB J, IEEE TKDE 等) 上共发表了 90 余篇学术论文。林学民教授累计在数据库领域重要的国际学术会议和国际学术期刊发表并录用了 200 余篇论文，其中 12 篇国际会议优秀论文。他在关键字查询方面的科研成果分别获得了国际顶级会议 ICDE'07 上的优秀学生论文奖 (Best Student Award) 和被列为国际顶级刊物 IEEE TKDE 2011 年 12 月份的 Spotlight Paper；在时空数据实时连续监控查询方面的科研成果被分别评为国际顶级会议 ICDE'10 和 ICED'12 上的优秀论文 (One of the Best Papers)；在文本数据质量管理方面的科研成果被评为国际顶级会议 SIGMOD'11 上的优秀论文 (One of the Best Papers)。他于 2010 年入选国家第四批“千人计划”。</p>	
<p>报告摘要：</p> <p>Graphs are widely used for modeling complex structured data, with a broad spectrum of applications such as biochemistry, bioinformatics, web search and social network, road network, etc. Over the last decade, tremendous research efforts have been devoted to many fundamental problems in managing and analyzing network data. Some interesting problems include graph structure search, node-pair similarity search, graph compression, uncertainty graph, and graph pattern matching.</p> <p>In this talk, I will provide an overview of recent advances in network data based search, and address some of the open issues. The technical part will mainly consist of two parts graph structure search and node-pair similarity search. More specifically, (1) graph structure search includes subgraph containment search, similarity subgraph search, supergraph containment search, similarity supergraph search, exact all-matching, and similarity all-matching. Most of these problems are challenging due to their NP-hardness, and they need to be resolved by developing efficient and effective techniques to cope with large network data. (2) Node-pair similarity search involves two promising link-based similarity measures SimRank and</p>	

SimFusion. Due to their self-referential concept, the sheer size of the Web has presented striking challenges to their fast computations. I will introduce novel techniques to efficiently and effectively compute node-pair similarities on large graphs.

新进展报告 2	海量 RDF 数据管理与应用
时间	2013 年 8 月 17 日 14:45-15:30
	杜小勇 中国人民大学
<p>个人简介： 杜小勇，中国人民大学教授，博士生导师，工学博士。现任中国人民大学信息学院院长，教育部数据工程与知识工程重点实验室主任。学术兼职包括教育部科技委信息学部委员，教育部文科计算机基础课程教指分委主任，国家自然科学基金委第信息科学部计算机科学专家评审组成员，中国计算机学会常务理事、专委工委主任、青年科学家奖评选委员会主席等。中国计算机学会大数据专家委员会委员，中国通讯学会大数据专家委员会委员。曾担任国家 863 计划软件专项专家组专家、数据库重大专项专家组组长、“核高基”重大专项实施方案编写专家组成员等。</p>	
<p>报告摘要： RDF 数据并不是一个新的概念，最初是作为语义网的组成部分被广为所知。本报告首先阐述了为什么数据库学术界开始逐渐重视 RDF 数据管理问题，之后介绍 RDF 数据管理的现状，以及我们在 RDF 数据管理方面所做的一些创新工作。</p>	

新进展报告 3	数据系统基准评测：回顾与展望
时间	2013 年 8 月 17 日 15:50-16:35
	周傲英 华东师范大学
<p>个人简介： 周傲英，华东师范大学长江特聘教授、云计算与大数据研究中心主任、海量计算研究所所长。他分别于 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。目前担任中国计算机学会数据库专业委员会副主任、ACM SIGMOD China 副主席。担任《计算机学报》副主编，《VLDB Journal》和《World Wide World Journal》等期刊的编委；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、大数据基准测试和性能优化。</p>	
<p>报告摘要： 数据库基准评测(Database Benchmark)在成就数据库产业近四十年的辉煌中具有不可替代的作用。基准评测相当于数据库系统中的技术标准和规范，有关基准评测的深入研究极大地推动了数据库管理系统的发展、成熟和应用。本报告通过回顾以 TPC 系列为代表的数据库基准评测的发展历程，从方法学层面来进行认识和总结，为“大数据”时</p>	

代的数据系统基准评测研究提供参考和借鉴。面对“大数据”，大一统的数据管理模式不再合适，它促使人们探索 RDBMS 以外的数据管理方式，就好像又回到数据库系统刚刚从文件系统蜕变而成的那个年代，未来会呈现出更加丰富多彩的局面。时下来看，Hadoop 是其中一个优秀代表。继之而来的是，针对一类应用会发展出一类特殊的数据管理技术和系统。“大数据”研究的典型特征是它与应用密切相关，对理解中国式应用我们有天然的优势。我们从事数据管理研究的科技人员处在一个充满机遇的时间窗口，开展有应用针对性的大数据基准评测是抓住这个机遇的重要环节。本报告将探讨如何通过基准评测研究来带动社交媒体、中国金融、位置服务等在我国具有广泛应用前景的数据管理技术和系统的发展。

新进展报告 4	大数据环境下数据库技术的变革-事务与并发控制
时间	2013 年 8 月 17 日 16:35-17:20
	<p style="text-align: center;">李战怀 西北工业大学</p>
<p>个人简介：</p> <p>李战怀，教授，博士生导师，西北工业大学软件技术研究所所长。中国计算机学会数据库专委会副主任委员；中国计算机学会信息存储专委会副主任委员；中国航天学会计算机应用专业委员会委员等。先后主持了国家 973 课题、国家自然科学基金重点项目、国家自然科学基金重大国际合作项目、国家 863 计划重大项目课题、国防科技预研、国家自然科学基金面上项目、国家 863 项目等。近年来，在数据库和相关领域公开发表的论著（译著）7 部，在有关国际和国内学术期刊和会议上发表论文 200 余篇。数十次担任国际（国内）学术会议的程序委员会委员（或主席），获省部级二等奖两项。主要研究领域：数据管理技术、海量信息存储技术等。</p>	
<p>报告摘要：</p> <p>数据库技术从 20 世纪 60 年代中期诞生以来发展极为迅速，被广泛的应用于制造业、金融、交通、能源、通讯、旅游、政府等众多领域，并获得了极大的成功。其中，事务处理和并发控制是关系数据库保证系统一致性的最重要的功能之一。进入 21 世纪后，在大数据时代背景下，应用需求、支撑架构、运行环境都发生了重大的改变，传统的事务概念已经不完全适用。在应用需求的驱动下，ACID 约束逐步被放松，CAP 理论的出现为非关系型数据库等运动提供了新的适用法则，不同的设计哲学触发了不同的技术变革，产生了不同应用环境下多样化的事务处理技术。本报告简要介绍了数据库事务相关的基本概念和技术发展脉络，分析了大数据环境对事务技术产生的影响变革、探讨了新型数据处理系统采纳的多种事务和并发控制技术，最后我们重新审视应用环境的变化，揭示了数据库技术领域特别是事务技术领域在大数据背景下的发展机遇与挑战。</p>	

企业报告

地点：启航剧场

时间：2013年8月18日 14:00-17:00

主持人：张志强、于旭

企业报告 1	下一代存储热点 - 软件定义存储
时间	2013年8月18日 14:00-14:30
	李丽虹 EMC 中国卓越研发集团高级软件经理
个人简介： 李丽虹 (Echo Li), EMC 中国卓越研发集团高级软件经理，分布式系统和网络存储技术领域的专家。领导团队负责软件定义存储以及云基础架构产品的研发。李丽虹毕业于哈尔滨工业大学理工实验班，机电一体化硕士学位	
报告摘要： 国际存储巨头 EMC 发布软件定义存储战略后，软件定义的数据中心，软件定义的网络，软件定义的服务器等概念正在业界热议中，2013 已然成为了“软件定义”元年。将借此机会解析软件定义存储并分享 EMC 的软件定义存储战略。	

企业报告 2	企业大数据管理与应用的技术挑战
时间	2013年8月18日 14:30-15:00
	冯是聪 北京秒针公司技术总监
个人简介： 冯是聪 2003年毕业于北京大学计算机系，获工学博士学位。作为核心成员之一，参与了第一代搜索引擎“北大天网”的研发。2003毕业后加入了朗讯(中国)贝尔实验室，担任高级软件工程师。2006年，加入刚刚创立的惠普中国研究院，担任高级研究员和团队带头人，主要负责企业搜索与云计算方向的研究工作。2011年加入北京秒针公司，担任技术总监，负责算法研究部。目前的主要研究方向包括计算广告学，大数据和云计算。	
报告摘要： 秒针公司 (http://www.miaozhen.com) 作为中国领先的第三方广告技术公司，目前日均处理数据超过 2TB，拥有日均处理 1000 亿广告请求的数据处理能力，累计存储、处理数据超过 2PB。目前超过 70% 的世界 500 强企业，包括宝洁、欧莱雅、强生、可口可乐、YUM、微软、大众、亚马孙等国际知名品牌都在使用秒针的产品和服务来优化其数字广告投入，提升广告投资回报率。本次报告首先将从算法和平台两个方面，跟大家分享	

秒针公司在管理和应用企业大数据过程中遇到的核心技术挑战：比如，跨屏数据打通、归因分析、各种广告定向技术、拍卖理论、广告转化率预测、多种数据源在统一云存储平台上的管理，等等。其次，还将分享秒针公司应用大数据技术服务大型企业（中国联通，中石油，北京电视台，国家统计局，等等）的经典案例。第三，将同大家分享 2013 年秒针公司举办的首届计算广告学大赛暨 RTB 算法大赛的情况。秒针公司通过公开海量的秒针数据集，来促进大数据以及计算广告学的学术研究以及优秀人才的培养。所以，最后将简要地介绍一下秒针公开数据集的情况。

企业报告 3	曙光大数据通用处理平台
时间	2013 年 8 月 18 日 15:00-15:30
	段刚 曙光信息产业股份有限公司总裁助理，教育行业事业部总经理
<p>个人简介：</p> <p>现任曙光信息产业股份有限公司总裁助理，教育行业事业部总经理，分管曙光公司教育行业全面销售与市场营销工作。</p> <p>多年来，段刚先生精耕于教育行业，在 IT 领域具备丰富的管理经验。作为中国 IT 产业发展见证者，段刚已将自己扎根民族企业——曙光公司，凭借自身丰富的教育行业 IT 市场管理经验，秉承“自主创新服务中国”企业发展理念，积极的推动着中国教育信息化进一步发展，助力中国教育的腾飞。</p>	
<p>报告摘要：</p> <p>1、通过对曙光公司的简单介绍，让大家了解曙光公司的历史、现状以及在云计算、大数据方面的未来发展蓝图；</p> <p>2、通过介绍曙光大数据解决方案与服务，阐述曙光大数据在平台能力及应用的灵活扩充和增长，提供真正的统一数据分析能力，并提供标准的平台接口，支持各行业丰富的第三方应用集成，包括智慧城市、金融、电信、交通、医疗、教育、军工等各行各业丰富的大数据应用软件，提供高效和个性化的数据分析服务，为企业提升工作效率、增强产业创新，为企业带来创新的技术模式、商业格局和投资机会，从而获得在信息时代的竞争优势。</p>	

企业报告 4	Social Information Filtering
时间	2013 年 8 月 18 日 15:50-16:30
	李航 华为诺亚方舟实验室首席科学家
<p>个人简介：</p> <p>Hang Li is chief scientist of the Noah's Ark Lab at Huawei. He is also adjunct professor of Peking University and Nanjing University. His research areas include information retrieval,</p>	

natural language processing, statistical machine learning, and data mining. He graduated from Kyoto University in 1988 and earned his PhD from the University of Tokyo in 1998. He worked at the NEC lab in Japan during 1991 and 2001, and Microsoft Research Asia during 2001 and 2012. He joined Huawei Technologies in 2012. Hang has more than 100 publications at top international journals and conferences, including SIGIR, WWW, WSDM, ACL, EMNLP, ICML, NIPS, and SIGKDD. He and his colleagues' papers received the SIGKDD'08 best application paper award, the SIGIR'08 best student paper award, and the ACL'12 best student paper award. Hang has also been working on the development of several products. These include Microsoft SQL Server 2005, Microsoft Office 2007 and Office 2010, Microsoft Live Search 2008, Microsoft Bing 2009 and Bing 2010. He has also been very active in the research communities and served or is serving the top conferences and journals. For example, in 2012, he is track co-chair of the web search track of WWW'12; senior program committee members or area chairs of WSDM'12, KDD'12, CIKM'12, ACML'12, AIRS'12; co-chair of KDD'12 summer school, etc; and an editorial board member on the Journal of the American Society for Information Science, ACM Transaction on Intelligent Systems and Technology, and the Journal of Computer Science & Technology.

报告摘要:

I will start my talk with an introduction to the Noah's Ark Lab of Huawei and our vision on big data. I will then introduce our ongoing work on social information filtering. The ultimate goal of our project is to construct information assistants for users and help them to easily and quickly access information. At the first step, we consider leveraging the information on social media such as Twitter and Weibo for construction of information assistants. We call such an approach "social information filtering". I will explain the basic problems of social information filtering, and discuss the technical challenges and potential solutions to the challenges. Finally, I will introduce our ongoing experiments on social information filtering at Weibo.

企业报告 5	SAP HANA 技术架构
时间	2013 年 8 月 18 日 16:30-17:00
	<p style="text-align: center;">黎文宪 SAP HANA 创新应用研发全球副总裁 SAP 中国研究院副总裁</p>
<p>个人简介:</p> <p>黎文宪博士现任 SAP 全球副总裁和中国区创新应用研发部的负责人。他负责搭建基于 HANA 内存数据库预测分析能力和相关应用, 及战略性合作项目, 大力支持中国区 HANA 伙伴生态系统、创业公司合作项目、客户合作以及云基础设施的构建, 以及主导与包括清华大学在内的国内知名高校的联合研究计划。黎博士在美国西北大学获得计算机科学博士学位, 并持有金融方向的工商管理硕士学位(MBA)。他的研究领域包括数据库、内存计算、数据挖掘、优化/调度以及开发新颖极致的应用。在加入 SAP 之前, 黎博士在位于美国硅谷的 IBM 阿尔马登研究中心(IBM Almaden Research</p>	

Center)任职, 并获 2007 年的 IBM 科研奖和 2004 年的 CeBIT 美洲最佳企业软件奖。他曾发表过覆盖广泛领域的 100 多份期刊论文和会议论文, 编辑了 Springer 出版的 3 本书籍, 并拥有超过 50 项已授权/申请中的美国专利。他同时活跃于国际学术会议, 担任 SIGMOD, ICDE, VLDB, WWW, ICDCS 等知名国际会议的程序委员会主席或程序委员。他并曾于美国 Rutgers University 任教。

报告摘要:

SAP HANA is an in-memory data platform that is deployable as an appliance, or in the cloud. At its core is an innovative in-memory relational database management system which makes full use of the capabilities of current hardware (RAM and multi-core CPUs) to increase application performance, to reduce cost of ownership and to enable new scenarios and applications that were not possible before. This speed-up processing capability from HANA could help existing applications to achieve the goals that would be considered impossible in the past, and also help to re-design and improve current business process flow in enterprise. Furthermore, we can come out new applications based on modern innovation ideas.

SAP HANA also extend it's traditional RDBMS capability to a hybrid data management system :

- Support multi in-memory storage : Row-based , Column-based , Object-based
- Not just only support traditional structure data management function (RDBMS) , but also provide un-structure data management such as Text Analysis , Nature Language Process , FullText indexing and Searching
- Support GIS application function.
- Support REST service provision function combined with OData & JSON (XS Engine). Allow cross-platform UI technology to connect with HANA platform directly without using other middle layer solution.
- Multi scale-out solution to fulfill all kinds of landscape requirements : Shared storage (NAS , NFS ...), Non-Shared storage , Data Center replication
- Support scale-out data partition mechanism which is used combined with multi-core CPUs by parallelization of execution to improve running performance
- Still use single SQL interface to support above multiple features
- Eliminate complexity of system landscapes , reduce TCO

In addition to introduce the topics above, I will also show some research prototypes done with HANA.

研究生学术辅导

地 点：八一宾馆多功能报告厅
时 间：2013年8月16日 13:30-17:10
主持人：韩启龙

学术辅导报告 1	Big social multimedia: opportunities and challenges
	Hengtao Shen, University of Queensland
<p>个人简介： Heng Tao is a Professor of Computer Science and an ARC Future Fellow in the School of Information Technology & Electrical Engineering (ITEE) at the University of Queensland (UQ). He obtained his BSc with 1st class Honours and PhD from Department of Computer Science, National University of Singapore (NUS) in 2000 and 2004 respectively. He then joined the University of Queensland as a Lecturer, Senior Lecturer, Reader, and became a Professor in late 2011. His research interests mainly include Multimedia/Mobile/Web Search, and Big Data Management on spatial, temporal, multimedia and social media databases. Heng Tao has extensively published and served on program committees in most prestigious international forums of multimedia and database areas. He received the Chris Wallace Award for outstanding Research Contribution in 2010 conferred by Computing Research and Education Association, Australasia. He is currently an Associate Editor of IEEE Transactions of Knowledge and Data Engineering, and will serve as a PC Co-Chair for ACM Multimedia 2015.</p>	
<p>报告摘要 Abstract: In the era of big data, the amount of social multimedia has reached an unprecedented level and keeps growing exponentially. In this talk, we will discuss the phenomenon of big social multimedia and its emerging opportunities and challenges in related research communities. In particular, I will introduce some recent progress in several promising directions, including automatic media tagging, near-duplicate utilization in social applications, and scalable indexing methods.</p>	

学术辅导报告 2	云计算中的可信查询处理: 一些研究经验分享
	徐建良, 香港浸会大学
<p>个人简介： 徐建良, 香港浸会大学计算机系副教授, IEEE、CCF 高级会员。主要从事移动计算、云计算等新型应用环境下的数据管理研究。近年来在相关领域共发表了 120 余篇</p>	

学术论文，其中 40 余篇长文发表在 CCF 推荐的 A 类期刊（TODS、TKDE、VLDBJ 等）及 A 类国际会议（SIGMOD、VLDB、ICDE 等）。根据 Google Scholar 的统计，目前 H-index 为 33，论文被引用次数超过 3200 次，单篇引用数超过 50 次的 22 篇。曾担任 IEEE ICDCS'12 TPC Vice-Chair，DASFAA'13 Demo Co-Chair 及 40 多个国际会议如 SIGMOD、VLDB、ICDE、WWW、DASFAA、WAIM、APWeb 及 NDBC 等的程序委员会委员，并获得 2010 年度 IEEE ICDE 会议的“Outstanding Reviewer Award”。更详细信息可访问：<http://www.comp.hkbu.edu.hk/~xujl/>。

报告摘要

数据安全是当前云计算研究中的热点，其中一个难点是如何同时保障查询结果的完整性和数据的隐私性。本次报告将分享我们在进行可信查询处理技术研究过程中的一些故事，并以此为例介绍在寻找新问题、解决新问题、撰写及发表论文等方面的一些经验。

学术辅导报告 3	浅谈科研工作与学术论文
	杨晓春，东北大学
<p>个人简介：</p> <p>杨晓春，东北大学教授，博士生导师。2001 年在东北大学获博士学位。2001 至 2003 年在美国 Brigham Young University 从事博士后研究。先后在美国 UC Irvine，香港中文大学和香港科技大学进行多次访问研究。主要研究兴趣包括数据库理论与技术，文本搜索与隐私保护技术。发表论文一百余篇，包括 SIGMOD，VLDB，ICDE 等顶级会议论文。获教育部自然科学二等奖、辽宁省科技进步三等奖。入选辽宁省百千万人才百人层次，获霍英东教育基金会优选课题资助，2006 年入选教育部新世纪优秀人才支持计划。</p>	
<p>报告摘要</p> <p>做科研与发表学术论文是相辅相成的。本报告与大家分享和交流我们近期在文本比对方面的研究经历，从科学研究、论文撰写、论文提交、论文报告等几个环节阐述科研与论文间的关系。</p>	

学术辅导报告 4	感知大数据的近似计算技术
	程思瑶，哈尔滨工业大学
<p>个人简介：</p> <p>程思瑶，生于 1982 年，黑龙江省人。于 2001 年考入哈尔滨工业大学，就读于计算机科学与技术学院，分别于 2005 年、2007 年在该学院获得工学学士、工学硕士学位，于 2012 年获得哈尔滨工业大学计算机软件与理论工学博士学位，师从李建中教授。2012 年 12 月至今在哈尔滨工业大学计算机科学与技术学院任讲师一职。程思瑶博士在读博期间和工作之后对无线传感网中的感知数据计算技术进行了较为系统深入的研究，率先开展了海量感知数据(ϵ, δ)-随机近似计算、物理世界的 $O(\epsilon)$近似等几项全新研究，并</p>	

在国际上取得了具有一定影响力的研究成果。目前,程博士发表和录用 14 篇学术论文,其中在计算机网络顶级国际会议 INFOCOM、在分布式系统顶级国际期刊 IEEE Transaction on Parallel Distributed Systems(TPDS), 在分布式系统一流国际会议 ICDCS 上发表和录用的论文 7 篇。并且,程博士作为主要成员先后参与了国家重点基础研究发展计划(973 计划)项目 1 项、国家自然科学基金重大项目 1 项、青年科学基金项目 2 项。此外,程思瑶博士还是 IEEE International Performance Computing and Communications Conference (IPCCC 2013)的程序委员会成员,国际顶级期刊与会议 IEEE Transactions on Knowledge and Data Engineering(TKDE)、INFOCOM 等特约审稿人。

报告摘要

随着传感网技术的迅速发展和广泛应用,感知数据的规模爆炸式地快速增长,目前许多系统年产感知数据的规模已超过了PB 量级,例如,国际气候研究组织在2010 年所获取的气象数据已超过2.5PB,并且正以指数速度增长,预计到2020 年将超过100PB;北京市目前安装了40 万个电子眼,每年产生1440PB监视数据等。Science 发表的论文指出,世界范围内的感知数据增长率为每年58%,而全世界的数据存储能力的增长率为每年40%,2007 全世界产生的感知数据已经超过全世界存储容量,2011 年全世界产生的感知数据已经超过全世界拥有的存储器总容量的二倍。显然,可以说“感知大数据”时代已经来临。与以往简单的感知数据不同,感知大数据呈现出“规模庞大”、“相关性强”、“质量不高”等诸多新特征,这使得感知数据计算方法的设计面临许多挑战。如果不解决这些挑战性问题,信息社会的发展将遭遇极大的障碍,亦将极大影响国民经济和社会的发展。本报告面向感知大数据,主要对我们目前在感知大数据近似计算方面所取得的研究成果进行介绍,包括:面向物理过程 $O(\epsilon)$ 可恢复的高质量感知大数据的获取技术及感知大数据的 (ϵ, δ) -近似计算技术;并且对未来感知大数据计算所面临的挑战及研究前景进行初步展望。

分组报告列表

特 别 提 示:

- 1.每篇论文的报告时间为 10 分钟，回答问题 5 分钟。
- 2.会场提供笔记本电脑，请报告人提前将 PPT 拷贝到电脑中。由于提供的电脑不一定支持 Office2007/2010，建议按 PowerPoint2003 格式或者 PDF 格式准备 PPT，以免出现播放问题。

分组报告 1

地 点：启航活动中心四楼

时 间：2013 年 8 月 18 日 8:30-10:00

分组 A1-查询处理与查询优化

主持人：王晓阳/邢春晓

- 1.基于相变存储器和闪存的数据库事务恢复模型
范玉雷, 孟小峰 (中国人民大学)
2. 外存中高效的字符串相似性查询处理
王金宝, 高宏, 李建中, 杨东华 (哈尔滨工业大学)
3. 基于替换概率的闪存数据库缓冲区替换算法
林子雨, 赖明星, 邹权, 薛永生, 杨思颖 (厦门大学)
4. 并行 OLTP 系统中增量数据的自动划分
WANG Xiaoyan, CHEN Jinchuan, DU Xiaoyong, FAN Xu (中国人民大学)
5. 一种基于信息论的数据导向模式匹配方法
赵晨露, 申德荣, 寇月, 聂铁铮 (东北大学)
6. 面向 BSP 模型的负载均衡 Hash 图数据划分
周爽, 鲍玉斌, 王志刚, 冷芳玲, 于戈, 邓超, 郭磊涛(东北大学)

分组 B1-数据挖掘和知识发现

主持人：乐嘉锦/王晓玲

1. 面向用户观点分析的多分类器集成和优化技术
林煜明, 朱涛, 王晓玲, 周傲英 (华东师范大学)
2. 混合的大规模数据库自动模式抽象
王雪, 周烜, 王珊 (中国人民大学)
3. 差分隐私保护下一种精确挖掘 Top-k 频繁模式方法
张啸剑, 王淼, 孟小峰 (中国人民大学)
4. 基于量子协同的二分图最大权完美匹配求解方法
印桂生, 崔晓晖, 董红斌, 董宇欣, 崔香 (哈尔滨工程大学)
5. 支持动态图数据的子图查询方法
王楠, 王斌, 李晓华, 杨晓春 (东北大学)
6. 高度可伸缩的稀疏矩阵乘法
吴志川, 毛琛, 韩蕾, 陈立军 (北京大学)

分组 C1-Web 数据管理与数据库

主持人：陈红/张岩

1. 一种基于情感符号的在线突发事件检测方法
张鲁民, 贾焰, 周斌, 赵金辉, 洪锋(国防科技大学)

2. 基于敏感属性值语义桶分组的 t-closeness 隐私模型
张健沛, 谢静, 杨静, 张冰 (哈尔滨工程大学)
3. 一种基于分析特征与动态步长的微博排序学习算法
周诗龙, 徐俊刚 (中国科学院大学)
4. 面向查询意图的信息检索技术研究
张志强, 彭晴晴, 谢晓芹, 冯晓宁 (哈尔滨工程大学)
5. 微博网站中面向主题的权威信息搜索技术研究
杨平, 王丹, 赵文兵 (北京工业大学)

分组 D1-时空数据库与数据质量控制

主持人: 唐常杰/元昌安

1. 基于签到数据的短时间体验式路线搜索
宋晓宇, 许鸿斐, 孙焕良, 刘俊岭 (沈阳建筑大学)
2. 融合语义特征的移动对象轨迹预测方法
黄健斌, 张盼盼, 皇甫学军, 孙鹤立 (西安电子科技大学)
3. 基于 iMeMex 数据模型的数据空间索引方法研究
王红滨, 周连科, 王念滨, 邓胜春 (哈尔滨工程大学)
4. CUU: 一种大规模时空数据区域查询与更新策略
高振龙, 李红燕, 苗高杉, 雷凯, 王腾蛟 (北京大学)
5. 防止路径攻击的加权社会网络匿名化技术
陈可, 刘向宇, 王斌, 张弘毅, 杨晓春 (东北大学)

分组报告 2

地 点: 启航活动中心四楼

时 间: 2013 年 8 月 18 日 10:20-11:35

分组 A2-查询处理与查询优化

主持人: 冯玲/高云君

1. 一种列存储数据仓库中的数据复用策略
王梅, 周娇玲, 乐嘉锦 (东华大学)
2. 一种扩展条件函数依赖的发现算法
刘显敏, 李建中 (哈尔滨工业大学)
3. PAA: 海量数据上一种有效的近似聚集查询算法
韩希先, 李建中, 高宏 (哈尔滨工业大学)
4. 大规模概率数据上基于 EMD 距离的并行 Top-k 相似性连接算法
雷斌, 许嘉, 谷峪, 于戈 (东北大学)
5. Pass-Join-K: 多分段匹配的相似性连接算法
余海洋, 林琛, 陈珂, 江弋, 邹权 (厦门大学)

分组 B2-数据挖掘和知识发现

主持人: 印鉴/陈群

1. 一种高效的属性图聚类方法
吴烨, 钟志农, 熊伟, 陈莹, 景宁 (国防科技大学)
2. 一种新颖的医学图像建模及相似性搜索方法

潘海为, 李鹏远, 韩启龙, 冯小宁, 谢晓芹(哈尔滨工程大学)

3. 障碍空间中保持位置隐私的最近邻查询方法

朱怀杰, 王佳英, 王斌, 杨晓春(东北大学)

4. 一种大规模图数据上已知项搜索的优化方法

钟鸣, 王盛, 刘梦赤(武汉大学)

5. 基于图模型的医学图像聚类算法

潘海为, 谷井子, 韩启龙, 谢晓芹, 张志强, 李鹏远(哈尔滨工程大学)

分组 C2-Web 数据管理与数据库

主持人: 崔斌/李金宝

1. 基于键规则的 XML 实体抽取方法

刘显敏, 李建中(哈尔滨工业大学)

2. 众包质量控制策略及评估算法研究

张志强, 逢居生, 谢晓芹, 周永(哈尔滨工程大学)

3. 企业搜索引擎个性化表示与结果排序算法研究

李贵林, 杨禹琪, 高星, 廖明宏(厦门大学)

4. INVF: 面向信息网络的可视化框架与算法

李洋涛, 李川, 唐常杰等(四川大学)

5. ACT-LDA: 集成话题发现、社区发现和影响力分析的概率模型

吴良, 黄威靖, 陈薇, 王腾蛟, 雷凯, 刘月琴(北京大学)

分组 D2-云计算与物联网数据管理

主持人: 岳丽华/李国辉

1. 基于帕累托最优的 DaaS 数据布局策略

张甜甜, 崔立真, 徐猛(山东大学)

2. 基于浮动车的快速交通拥堵监控

吴佩莉, 刘奎恩, 谭毓安(北京理工大学)

3. 基于区域划分的数据流子空间聚类方法

于翔, 印桂生, 许宪东, 王建伟(黑龙江工程学院)

4. 混合云存储中海洋大数据迁移算法的研究

黄冬梅, 杜艳玲, 贺琪(上海海洋大学)

5. 云环境中海量数据的并行分组密码体制研究

师金钢, 郑艳, 孙焕良, 栾方军(沈阳建筑大学)

系统演示列表

地 点：启航阳光大厅

时 间：2013 年 8 月 17 日 13:30-17:00

注意事项：每个系统演示请准备一张海报（高 110cm*宽 90cm），于演示前粘贴到演示会场指定位置。系统演示不需要做口头报告。

系统演示

主持人：

- 1.HaoLap-基于 Hadoop 的海量数据 OLAP 系统
郭朝鹏, 王智, 韩峰 等, (东北大学)
- 2.基于 D-Ocean 的金融风险监控系统
冯普超, 邵春翡, 黄中东 等 (浙江大学)
- 3.SmartMS: Android 平台下灵活的海监执法取证及分析系统
黄冬梅, 张喆昱, 赵丹枫 等 (上海海洋大学)
- 4.TPCHSuite_一个 TPC-H 自动化测试工具的设计与实现
康强强, 江舟, 金澈清 等 (华东师范大学)
- 5.MMDVis:一个基于微博用户的多博文传播分析及可视化系统
郝竞超, 王朝坤, 司德谭 等 (清华大学)
- 6.基于众包的电子商务数据实体分类系统
叶晨, 王宏志, 周小田 等 (哈尔滨工业大学)
- 7.生物实验信息管理系统
王黎维, 付祖发, 刘东明 等 (武汉大学)
- 8.一种分布式系统上的元数据管理系统
魏光辉, 李杰斌, 王程玉 等 (复旦大学)
- 9.XML 数据上支持查询扩展的关键字检索系统
胡昊, 王君伟, 常橙 等 (复旦大学)
- 10.基于 RkNN 的村镇基础设施选址分析系统
许景科, 于戈, 孙焕良 等 (沈阳建筑大学)
- 11.CloudIndexEval:基于云平台的多维索引评测系统
周新, 王延昊, 刘辰 等 (中国人民大学)
- 12.TPCDroid: 面向移动智能终端的数据库查询能耗测试
杨濮源, 金培权, 岳丽华 (中国科学技术大学)

