



主持人郑伟民教授及会场

# 第一届大数据科学与工程国际会议(2016) 精彩荟萃

大数据时代的悄然来临，正在深刻变革着人们的科研、生产、生活方式，推动社会经济的发展、相关领域技术的创新发展，甚至国家竞争力的提升。大数据已经成为全球关注的热点，成为国内外学术界、产业界普遍关注的热点研究领域，在全球掀起了一个空前的研究热潮。

5月25日，第一届大数据科学与工程国际会议在贵州省贵阳市开幕。本届会议主要关注全球大数据发展趋势以及大数据在学科交叉领域中的研究应用，探讨大数据发展中面临的挑战、技术创新、应用创新等方向，着力反映国际大数据技术研究的最新进展。大会以“认识、科学、创新、人才”为关键词，旨在“以历史、发展的眼光认识大数据”“以合理、专业的视角解读大数据”“以跨界、融合的理念驱动大数据”“以开创、继承的思维培育大数据”。大会邀请了多位院士和国内外大数据领域的顶级专家学者作大会主论坛和各主题论坛的特邀报告。

## 主论坛：“大数据科学认识与理解”

5月25日上午，第一届大数据科学与工程国际会议（2016）的主论坛“大数据科学认识与理解”在贵阳金阳万丽酒店3F贵阳大厅2/3举行，清华大学教授、原中国计算机学会理事长、《大数据》编委会主任郑纬民作为大会主席及主论坛主持人，现场观众爆满，气氛热烈。



大会主论坛观众席，  
座无虚席



贵阳市政协副主席孙袁代表  
主办方出席并致辞



中国工信出版集团董事长，  
人民邮电出版社社长代表主办方  
出席并致辞



中国计算机学会大数据  
专家委员会副主任委员  
张晓东教授代表主办方致辞



## 郑纬民： 大数据发展需要产学研 结合，中国企业在大数据 领域有极大的发展空间

作为本次大会的主席，清华大学教授、原中国计算机学会理事长郑纬民表示，贵阳数博会是全球大数据领域的盛会，李克强总理的出席更说明了国家对于大数据产业的重视。第一届大数据科学与工程国际会议是大数据学术领域的具有世界一流水平的会议，作为数博会唯一的学术活动，它的举办也是希望对于开始蓬勃发展的的大数据领域，能够实现产学研的无缝结合。本次主论坛以“大数据的科学认识与理解”为主题，希望大家能够正确认识和理解大数据及其价值，下午设置的4个主题论坛，分别从大数据分析与管理、大数据系统、大数据安全、大数据学科与人才4个维度探讨大数据的系统性研究及人才领域的建设。

大数据的发展已经有几年的时间，大数据已经成为大势所趋，这已成为无可争议的事实。总的来看，大数据的商业变现仍然不多，他认为这需要一个过程。大数据是一个新兴领域，中国企业在大数据领域仍有很大的机



郑纬民教授接受采访

会。首先，目前国际大数据公司在大数据领域的开发均是开源的，公开源代码，中国企业可以直接获益。其次，中国是第二大经济实体，有大量的数据积累，中国在大数据的发展上有很大的希望取得突破。



## 俞士纶： 异构将让大数据实现更 多价值

美国伊利诺伊大学芝加哥分校特聘主任教授、清华

大学数据科学研究院院长，ACM/IEEE Fellow 俞士纶作

为大会的名誉主席在主论坛上发表了“对异构数据源的思考”的演讲。他认为：大数据时代，数据可以改变很多细节，我们要做的是找到有意义的数据，并融合不同的数据，将数据运用在实际的事情上。

俞士纶教授将数据库比作药品，将数据比作药品的内部结构：如果我们把来自不同领域的数据库和数据联系在一起，我们将可以利用数据做出更好的预测。

俞士纶教授给出了相似的案例，豆瓣利用人们对于一个电影的喜爱程度，综合这部电影和这个人以往看过电影的主演、导演、电影类型、人物性格等多种电影资讯就可以进行预测，抓住人与人之间的联系，做出预测。但这种数据预测的效果又是因人而异的，有些人社交关系比较重要，但有些人却没有朋友，所以数据的预测要

通过不同的路径进行预测，就会减少误差。

他认为：异构数据可能是一张图和与之相关的一段图解，这是以两个不同的方式来表达一件事。当我们从图和文字找到相似的地方时，通过图就可以找到相关的文字描述，通过文字描述就可以找到相关的图片。如果将两者联系在一起，结果能更为准确。

俞士纶教授表示：现在是大数据时代，重要的事情是我们要怎么融合不同的数据。比如气象预测，光做预测是根本赚不了钱的，但如果应用在货物运输、农业等具体领域，这些天气数据可以改变很多细节，提供更多便利，成为真正有价值的信息。所以，我们首先要了解什么是有意义的信息，然后考虑如何把这些信息合在一起去提供更好的服务。



## 邬贺铨： 大数据“博大精深”， 要强调规律，更深入运用

中国工程院院士邬贺铨作为大会的名誉主席在会上作了“大数据之博大精深”的演讲，他表示，目前已有的数据量相当大，具有“博”“大”“精”“深”的特点。

“博”指数据来源非常丰富，包括社会空间数据、物理空间数据和信息空间数据。按网络层面划分则包括自媒体数据、日志数据、富媒体数据、基础网络数据。其中国家安全数据、商业秘密数据、个人隐私数据，这3方面的数据不能开放，开放的应是其他的数据。

“大”指现有数据量已经远远超过传统的研究范围。现在互联网的数据流量每年新增40%，每两年的数据量就增加了两翻，可以预计，到2020年，全球的数据量

要超过40 ZB。大数据需要的数据量大，当数据量较少的时候，采用矩阵聚类分析的误差较大，只有当数据大于一定量的时候才可以使用。

“精”不是指追求精准，还是指强调规律。这可以理解为一个大数据集而言，我们很难要求所有数据都精准，只要大多数数据有代表性，并不妨碍数据集得到较为准确的结果。

“深”指对大数据的运用应该更深层次。大数据的分析不仅需要大量的数据，还需要建立模型。现在利用GPU可模拟超大型人工神经网络，最近百度、谷歌在这方面都取得了较好的进展。



## 方滨兴： 云监控与云加密打造 可信云

中国工程院院士方滨兴在会议上表示，云安全分为4个层面：第一是可靠的云，能够始终可靠地提供服务；第二是安全云，保障用户及租户在云中不会因攻击而受到损害；第三是可信云，保障云服务商不会对租户的运行进行侵害；第四是可控的云，保障租户不会利用云运算恶意程序。

大数据由云平台进行存储和处理，分工更细，另外，数据的所有权和管理权分离，这种分离导致数据不可控。而云监控与云加密可打造可信的云，云监控可防止云

服务商作恶，保障云中数据安全；云加密可降低泄密风险，无缝数据加密传输、支持任意云端平台。

方滨兴院士表示，云监控与云加密有七大特点：第一，增加了云管控与审计，具备对云平台和虚拟机全生命周期的细粒度管控和审计能力，具备对云平台和虚拟机的安全加固能力；第二，多终端无缝介入，随时进行操作；第三，智能权限策略，用户进行敏感操作，可以拒绝；第四，精准指令控制；第五，文化内容审计；第六，大数据分析；第七，更高安全系统。



## 梅宏： 我国的大数据发展应 回归理性

中国科学院院士、上海交通大学副院长、IEEE Fellow梅宏在会上做了精彩的报告。梅宏院士认为：大数据作为一个现象的产生是由于信息技术不断降低的成本以及互联网及其延伸带来的无处不在的信息技术的广泛应用。世界各国高度重视大数据，但是存在几个问题：大数据的应用还属于初级发展阶段，现在很多数据、案例很难被看作典型的成功案例，很多是传统的方法在技术上的拓展——“新瓶装旧酒”，属于统计的范畴。此外技术支撑能力也很不足。

梅宏院士表示：我国的大数据发展应回归理性。大数据对社会产生了重要的影响，但不能把所有的事都依靠大数据来解决。它可能会对我们的生产、生活带来负面影响，如可能带来数据所有权、隐私保护及信息安全等一系列的问题。当前的道德建设、法律建设以及技术手段建设仍处于滞后状态，而投资已经很热，应该顶层规划、示范引导，要积极谋划，审慎推进，避免一哄而上，造成超前投资或重复投资，总之，应该“利用驱动”、“需求驱动”。



## 高光荣： 发展大数据必须突破 三个门槛

美国特拉华大学电子与计算机工程学院终身教授、ETI公司创始人与首席科学家、ACM/IEEE Fellow高光荣认为，发展大数据要突破3个门槛：其一，认清大数据和高性能计算的依托；其二，找到把大数据变成知识的途径；其三，计算出大数据投资后产生的经济效益。

高光荣教授说，自己在国外一直从事高性能计算领域的研究。2005年，他接到一个项目，要求分析数据速度特别快。为了解决这个问题，他和团队合作制造出一个机器——世界上第一台大数据的超规模计算机，这些经验让他对大数据有了直观的认识。然而，高性能并非计算的“又一春”，眼下的高性能计算和大数据产业既面临挑战，也拥有机遇。

高光荣教授认为，为了突破上面3个发展门槛，行业专家们需要先把自己的“门槛”打开，互相学习，看看大



高光荣教授接受采访

数据和高性能计算到底有多少东西是共通的，从而找到维持大数据发展的办法。



## 华云生： 打破大数据研究、 应用障碍

香港中文大学常务副校长、伟伦计算器科学与工程学讲座教授、ACM/IEEE/AAAS Fellow华云生做了题为

“大数据研究的战略思考”的精彩演讲。华云生教授提出众多实例，说明在对大数据进行研究的过程中，目前还面

面临着很多挑战,包括对有用数据甚至是大数据的获取、对核数据的获取等。

华云生教授认为,应该打破现存的这些障碍,并列举了当前国际上对大数据进行的研究项目,提出了大数据研究的领域,包括环境、安全、运输、教育、通信、网络安全、智慧城市、金融、医疗等。华云生教授最后说,大数据发展需要有一个需求,有了需求才可以做出结果。另外,对大数据我们要从上到下进行管理,积极打破障碍才可以做出有用的大数据研究。



华云生教授接受采访



## 张晓东： 数据是检验真理的 重要标准

美国俄亥俄州立大学教授、ACM/IEEE Fellow张晓东在会上就“数据分析将对人类自身的了解达到前所未有的高度”做了精彩的报告。

张晓东教授表示,随着计算机的不断演进、数据量逐渐增大,数据处理面临诸多挑战。第一个挑战是原有的通用数据库技术不能满足大数据可扩展和高容错的需求,第二个挑战是传统的数据管理的商业模式不被日益增加的非盈利用户接受。大数据分析是走向理性决策的重要基础。大数据分析主要目的是学到知识,从巨大、不规则和零散的数据中找到新知识,既可以做出完美科学决策,同时可以消除感性、教条和口号式的决策,也可以凭借经验和直觉将其作为重要的参考。大数据让我们更清晰地看到社会本质和认识自身,从而使违背规律和人的本性的决策大大减少。大数据分析的本质不是数据量的大,而是分析结果的科学性。

张晓东教授感慨道:20世纪90年代时,预测互联网的出现和完备将对整个工业生产和人类日常生活做



张晓东教授接受采访

出颠覆性的变化,现已全部应验。今天通过对大数据分析,我们对人类自身的了解将达到一个前所未有的高度,越来越多的普适价值是以数据为基础的。数据是检验真理的一个重要标准。

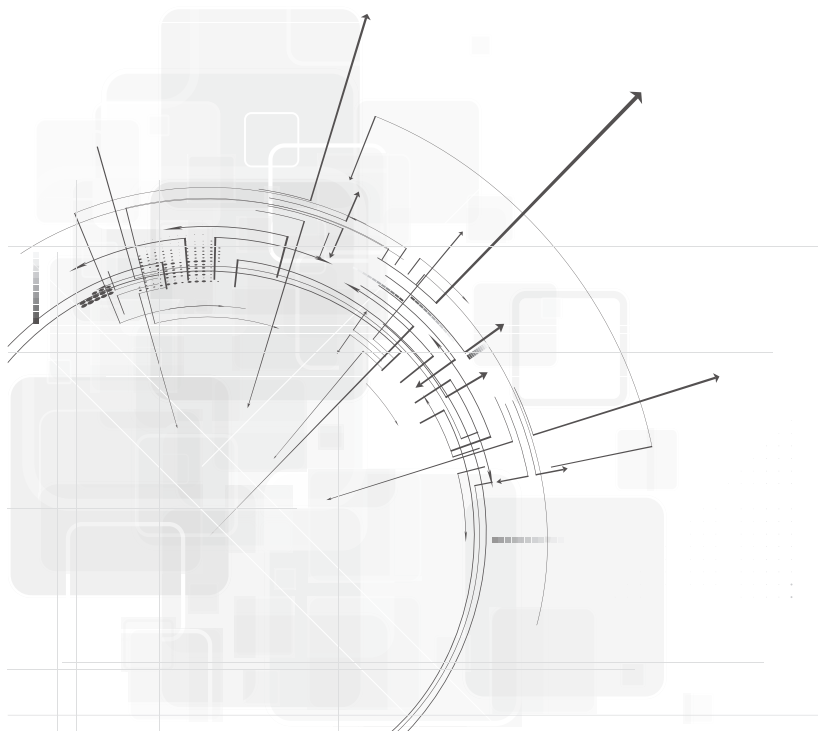


## 郭毅可： 数据科学是当代科学的 重要基础和方法论

英国帝国理工学院数据科学研究所终生教授、所长郭毅可就“数据科学和科学数据”做了精彩报告。

郭毅可教授介绍了英国帝国理工学院数据科学研究所主要的工作方法和思路。他认为大数据存在的原因，不是计算机发展太快，而是科学数据发展太快。英国帝国理工学院数据科学研究所主要做数据驱动科学方面的研究，以数据科学作为技术来推动科学的交叉。数据科学研究所的主要任务包括5方面：一是做研究；二是做教

育，把数据科学内容嵌入各个系的硕士教育中心；三是开发一些技术，为学校科学研究提供帮助；四是与工业合作，获得比较多的科研经费；五是影响力，跟政策、政府、国际合作。关于数据研究，研究方向包括3个方面：一是继承与承接；二是感知与交互；三是学习与认识的方向。郭毅可教授认为数据科学是当代科学的重要基础和方法论，而且科学研究推动了数据科学的许多方面和基础方法的发展，他们将在这些方向上努力。



## 主题论坛：大数据分析与管理

5月25日下午，“大数据分析与管理”主题论坛在贵阳金阳万丽酒店3F贵阳大厅1举行，中国人民大学信息学院院长杜小勇教授担任论坛主席及主持人。会上，嘉宾演讲精彩绝伦，现场气氛热烈，听众爆棚。会议开始，论坛主席杜小勇教授首先给观众解释了“大数据分析与管理”论坛主题的范畴，要考虑数据的采集、数据的存储、数据的分析、数据的管理，特别是数据的管理，因为任何一个技术的产业化，一定要考虑它的经济成本，要考虑它的经济价值，大数据同样也是如此。



“大数据分析与管理”主题论坛观众席



“大数据分析与管理”主题论坛讨论环节



## 杜小勇： 政府做好大数据治理 必须具备互联网思维

会议期间，中国人民大学信息学院院长杜小勇教授接受了数据观记者的采访，他表示：大数据正在成为这个世界上最重要的“土壤”和“基础”，成为一切管理和决策的依据。运用新理念、新技术、新方法对大数据进行全生命周期的创新管理和应用，是推动国民经济转型和社会管理创新的重要契机，也是提升国家综合竞争力的重要趋势。

他认为，大数据因其规模巨大、类型复杂、产生速度快、价值密度低等因素，对现有信息技术产生巨大挑战，需要运用新理念、新技术、新方法对其进行全生命周期

的创新管理和应用，从而促进国民经济的转型升级、社会管理的模式更新，乃至国家综合竞争力的全面提升。在大数据时代，有志于引领组织实现大数据转型的管理者们首先要养成看“数据怎么说”的习惯，重视数据。其次，要允许数据做主，放手让“大数据说话”，这将是改变组织决策文化的最大力量。

杜小勇教授表示，政府要做好大数据治理，必须深入理解互联网思维，这其中包括关联思维、去中心思维和人本思维。



## 裴健： 以大数据分析促进业务 观念和-content创新

加拿大西蒙弗雷泽大学教授、IEEE TKDE学报主编、ACM/IEEE Fellow 裴健在“大数据分析与管理”论坛作了题为“以大数据分析促进业务观念和-content创新的演讲。

裴健教授认为大数据和大数据的分析管理，不仅仅是技术问题，更多的是业务问题，跟业务开发和发展息息相关。从技术的层面出发，要提供数据管理到分析再到业务解决方案的全盘考虑。

他分享了其团队在大数据应用上面的认识和经验：首

先可以用大数据来承载灵巧的新兴应用，这是传统的应用领域与新兴的大数据应用领域的一个很大区别。通过创造性地使用丰富的数据，应用可以做得很小，开发快捷，而产生的社会效益却可能很大。其次，希望做若干应用能形成关键的业务流程。如通过现有工具进行欺诈检测时，很多工具会告诉我们这是疑似的欺诈案例，但鲜有系统直接支持我们怎样去调查一个案例，怎么分配人员和开销等。再次，要集中力量把业务的临界量超越过去，迅速推进多种相关业务，形成大数据支撑的新的业务群，使大家

能很快看到大数据带来的益处，才能够吸引更多的业务投入。最后，业务手段和内容创新实际上是相辅相成的。大

数据分析的作用体现在能否直接支撑业务部门进行业务手段和业务内容的创新。



## 黄锦辉： 大数据分析不一定可靠

香港中文大学工程学院教授、副院长黄锦辉在“大数据分析与管理”主题论坛作了精彩演讲。黄锦辉教授在其报告中，逆向看待大数据分析，对大数据分析的可靠性提出了质疑。他认为如果数据量不足，或数据的质量较差，或者数据是动态持续变化，那么分析的结果不一定是好的。另外，如果只靠商界人士开放数据，而政府不开放数据的话，出现“马太效应”的

可能性是很大的。

黄锦辉教授还提到，大数据是把双刃剑，要避免一些不良商人利用一些数据，做一些对消费者不利的行为，这是忽略消费者权益的事情，政府或者管理层应该把这点考虑清楚，需要制定法律监管、保护消费者的权利。大数据分析的关键是要有优质的数据、优良的模型算法。



## 周傲英： 应用驱动大数据创新

华东师范大学研究生院院长、数据科学与工程研究院院长、长江学者特聘教授周傲英在论坛上作了题为“从管理和分析说起：感悟大数据”的精彩演讲。

在大数据时代，对实时分析的需求很多。需要实时分析的问题有相应时间的要求，且数据量很大，需要分析的数据可随时注入，大数据的核心就是将这些数据关联起来，实时分析，实时注入，有弹性。大数据分析中事务

处理非常重要，需要高可靠的、高弹性的事务处理系统，支持现象级的应用。

周傲英教授表示，IT的发展，要立足应用，现在要去IOE，要讲安全可靠，自主可控。应用驱动创新，一方面要立足应用，另一方面需要开放，解放思想。如果中国开源环境好了，中国的创新一定会强起来的，万众创新，草根创新。



## 王腾蛟： 用工程的、科学的方法处理大数据才能得到正确的结论

北京大学文理大数据研究中心常务副主任，北京大学信息科学技术学院教授王腾蛟在论坛上作了精彩报告，题目为“大数据分析是一种科研方法吗”王腾蛟说《Nature》和《Science》，甚至《光明日报》都认为人文社会科学的大数据研究会成为未来人文社会科学的主要领域，他们认为这种方向性是没有问题的。当然反对方也有很多，但是仔细看看这些反对方的态度，不是否定大数据的方向，而是认为很多团队有了海量的数据之后，肯定都能得出一些结论，但是同样面对这样的数据，不同的

方法可能出来的结论是矛盾的，所以怎么用工程的、科学的方法去处理数据，得到正确的结论，这个是大家比较关心的。

王腾蛟教授认为，有了数据拿来分析可以得出一个结论，但是其正确与否需要用工程技术的方法作保证，大家要沉下心做工程分析，然后评估数据从收集到应用的技术环节。从工程技术保障上来说，个人很难做这样的事情，要做这样大的事情，在应用上有价值的事情，真的需要大家合作来实现共赢。



## 冯是聪： 大规模复杂关联数据在公安领域的应用

明略数据CTO冯是聪在“大数据分析与管理”主题论坛作了题为“大规模复杂关联数据在公安领域的应用”的精彩报告。冯是聪表示，明略数据是专门做大数据的，认为公安领域特别适合作为大数据的应用场景。他列举了3个犯罪实例证明大数据在公安领域会发挥巨大作用。用大数据或者大情报帮助公安部门抓犯罪分子的问题如下，第一个问题就是如何把结构化和非结构化的海量数据聚在一起；第二个问题是如何在危害性非常强的涉恐案件发

生前做预警；第三个问题就是时效性，时效性特别重要。

冯是聪用举例、视频演示的方式给出了大数据分析在公安领域的应用。最后就大数据分析技术应用在案例中的情况，提出了4个问题寻求领域专家的建议，包括：第一个是数据仿真的问题；第二个是预警问题；第三个是如何找一个评判模式，在大社群里面找出特异的样本；第四个是在超大规模上支持实时性和事务性，这和大数据里面很多理论是相悖论的。

## 主题论坛：大数据系统

5月25日下午，“大数据系统”主题论坛在贵阳金阳万丽酒店3F贵阳大厅2举行，论坛主席为中国计算机学会大数据专家委员会秘书长、中国科学院计算技术研究所副总工兼所长助理程学旗研究员，主持人为中国科学院计算技术研究所靳小龙副研究员。会场座无虚席，现场观众反响热烈。程学旗研究员表示，“大数据系统”论坛主要从两方面来进行讨论交流：一方面是如何使用大数据关键词，另一方面是如何把数据存储好，管理好，使用好。



“大数据系统”主题论坛会场



## 程学旗： “快数据”使大数据商业 价值最大化

程学旗研究员在会上做了“大数据引擎与分析系统”的精彩报告。程学旗研究员认为，大数据已从概念理解阶段到了价值挖掘阶段，需要大数据引擎来推动其发展。大数据引擎是数据从自然的分布存在到价值业务空间的“心脏”，分很多方面，包括数据的存储管理、数据的分布式计算以及对数据价值的挖掘等。

程学旗研究员表示，大数据的商业价值是企业的核心价值，对它的认识要从大数据分析流水线来看，不能孤立在这个点上。挖掘数据的价值，体量、精准度、时效性都是关键因素，未来大数据要与“快数据”结合起来，因为数据的商业价值会随数据分析时间增长而降低。在时效性上，流式数据处理具有优势，应以其为核心追求数据商业价值的最大化。同时，他表示，大数据的深度分析、分布式计算



程学旗研究员接受采访

框架、异质数据的管理将是大数据生态追求的3个方向。



## 陈文光： 以“性能优先”为目标 设计大数据系统

清华大学计算机科学与技术系教授陈文光在论坛上作了题为“大数据分析平台——从扩展性到性能”的精彩演讲。陈文光教授说，设计大数据分析系统最关注的两个因素是性能和扩展性，而现有系统中，这两个因素

是一对矛盾体，提高性能的时候必然减少容错性。这是错误的。可以使用更少的节点和更短的运行时间来完成同样的大数据分析任务，未完成容错需求甚至可以采用开销更大的容错技术，但即使这样，用于容错的时间仍

然要比现有系统小很多。我们应该追求比较好的性能和一定能力的容错,以性能优先来设计大数据系统,尽量使用较少的节点,可以通过有序访问节点进行控制、尽量减

少写入硬盘、设计结构优化等多种方式来实现,这是未来大数据所需要的。陈文光教授表示他们在做的系统就是基于这些思想,并已经初步通过测试进行了验证。



## 钱岭： 从点状应用走向大数据平台

中国移动苏州研发中心总经理助理兼CTO钱岭在“大数据系统”主题论坛结合中国移动所做的对大数据系统整合的实践,作了精彩演讲。

钱岭表示,运营商是数据资源优势拥有者之一,拥有“数据金矿”的运营商如何挖掘数据价值是其面临挑战。运营商希望建立一个企业级的大数据平台来满足通用的、定期的以及远期的需求。这样的企业大数据平台的功能要把资源、服务、数据、应用和用户统一关联起来,并提供统一的应用开发和运行环境。

现有运营商运营系统存在种类繁多、相互孤立、数据

分散在各个维度、建设成本高、管理开销大、安全风险高等问题。要整合这些系统建立大数据平台,首先要统筹考虑管理、业务重构和平台技术实现3方面的要求。其次针对数据源分散管理,需要解决4个问题,第一是企业管理问题,即解决体系架构调整的问题;第二是数据质量管理问题;第三是数据变现的管理问题;第四是数据安全的管理问题。再次是实现运营优化,通过建立技术体系,把平台管理、应用开发、商品解耦结合起来,明确技术体系架构,包括技术体系的基础来源,是自主研发还是采购,是集成还是开源等。



## 陈世敏： 新一代的NVM技术将给 存储带来崭新变化

中国科学院计算技术研究所研究员陈世敏在“大数据系统”主题论坛作了题为“非易失存储的大数据管理系

统优化”的精彩演讲。

陈世敏研究员表示,flash的发展主要朝两个方向。

一个是增加存储单位可存储比特数。另一个是以三维体表达更高的密钥，目前其发展受到了向下扩展难的制约。新一代NVM技术因此产生，其中发展最成熟的具有代表性的技术是PCM，另一个是STT-RAM（自旋扭矩转换存储器）。它们的共性是不存储电荷，而是改变单元电阻，最终结果是非易失的，是可字节寻址的，而且访问速度接近DRAM。所以在系统中，NVM可能会用于DRAM的替代和扩展。对内存出现故障的系统也可以提供非易失的存储，进行快速的恢复。当NVM系统真正替代DRAM系统时，它将会给存储系统带来崭新的变化，在现实中产生巨大的作用。



陈世敏研究员接受采访



## 刘睿民： 在流数据处理中要注意时效性

柏睿数据总裁兼CTO刘睿民在“大数据系统”主题论坛作了题为“基于流数据库及内存分析建立物联网数据应用”的演讲。刘睿民认为，流处理成为热门话题，是因为大数据的价值最大化实现基于数据的鲜活性，而流处理能够实现实时的数据分析处理。SQL数据库具有很多分布式的优化策略，在实时分析的流处理下，可以通过SQL来做表达，SQL的扩展性应对实时大数据的分析是非常有优势的。

通过SQL建立IoT应用，无论是不同的传感器还是建立模型，都可以通过ROW处理持续不断地查询、分析，不过必须要保证时效性，为此甚至需要重新审视忽视的细节。



刘睿民总裁接受采访

## 主题论坛：大数据安全

5月25日下午，“大数据安全”主题论坛在贵阳金阳万丽酒店3F贵阳大厅3举行。论坛主席中国科学院信息工程研究所副总工程师李凤华研究员主持了会议，各位嘉宾分析讨论了当前环境下大数据发展过程中存在的安全问题，令人警醒。



会场观众



## 李凤华： 大数据隐私保护亟待 新理念——隐私计算

中国科学院信息工程研究所副总工程师李凤华研究员在“大数据安全”主题论坛作了题为“隐私计算：大数据环境下的隐私保护”的演讲。

李凤华研究员表示，大数据发展面临规模、性能、安全、能耗几方面的挑战，需要以绿色节能为目标，形成兼顾大规模、高效率的面向大数据的整体解决方案。把握好大数据安全与隐私保护才有未来更加快捷、方便的服务。在大数据安全方面面临3个技术挑战：大数据中的多密级

数据机密性管理与使用技术；多隐私保护方案融合技术；大数据的数据机密性与隐私保护协同工作问题。

李凤华研究员提出，现有隐私保护方案虽然很多，但针对性强，很难形成融合架构，需要面向大数据隐私保护的新理念——隐私计算。隐私计算是泛在网络空间隐私信息保护的重要理论基础，是面向隐私信息权声明周期保护的计算理论和方法，也是高效能多系统融合的隐私信息保护，通过隐私计算可以充分发挥现有隐私技术各自的优势。



## 任奎： 模糊搜索的有效性更大

美国纽约州立大学布法罗分校电子和计算机工程系教授，IEEE Fellow任奎在“大数据安全”主题论坛作了题为“大数据安全搜索：挑战与进展”的演讲。

任奎教授首先介绍了数据搜索加密的现状，现有的搜索加密工具有两种，一种具备很好的安全性能，但带来大量的计算量，另外一种简单方便，但实际却存在很大安全漏洞。因此搜索加密工具是目前大数据安全检

索的挑战之一。另外一个挑战是模糊搜索。数据搜索应该支持所有的模糊搜索，模糊搜索采用比较的文本集，按照框架进行一系列的搜索，搜索得到的结果要根据与关键词的相关度进行排序，这样才能体现出模糊搜索方法的有效性。任奎教授最后提到，更广泛意义上来讲，模糊搜索也包括相似搜索，此时面对的不仅仅是文本数据，还可能是照片等其他类型的数据。



## 李晖： 云计算的用户隐私问题已 已经到了必须解决的阶段

西安电子科技大学网络与信息安全学院教授、执行院长李晖在论坛上作了题为“云计算环境下的数据安全技术”的演讲。

李晖教授认为，云计算的用户数据隐私问题已经到了必须解决的阶段，密码机制可以在一定程度上解决问题，但仍有许多待完善之处，如密码使用的方便性、功能的多样性、保证效率。如目前的数据加密机制从实用性考虑，基于属性加密只能用于加密密钥，大文件加密必须采用KEM机制。在密文搜索上，未来要在保证搜索功能实用化的同时，考虑4方面的问题：如何进一步提高搜索速度；什么样的安全性定义更符合实际需求；

安全性和实用性如何取舍；是否可以用一种新的设计思路——将搜索看成是一种外包运算。在数据完整性的审计上要考虑谁来充当TPA，能否进一步减少数据所有者的运算量，如何处理（算法和效率）包括不同用户的挑战和应答。由于用户在使用数据服务的工程中，不再是真正意义上拥有数据，所以如何保证存储在云端、不再需要的隐私数据能够安全销毁成为新的难点问题。在数据的安全销毁上，需要考虑几个关键点，谁来保存密钥，如何保证密钥被安全删除，如何安全存放密钥，如何保证加密数据被删除（因为随着时间的推移，加密算法会被破除）。



## 张卫明： 隐藏是应对隐私危机的一 一种思路

中国科学技术大学网络信息安全系（筹）副教授、副主任张卫明在论坛作了题为“信息隐藏与数据隐私保护”的演讲。

张卫明副教授表示，对于互联网、大数据时代，我们要关注其带来的忧患，也要换角度思考其带来的需求与价值，对于这个时代下大家关注的隐私泄露问题也同样如此，可以以隐私泄露—忧患与危机—需求—价值的思路

来考虑这个问题。如果有方法就可以解决隐私泄露问题，有思想就可以创造其价值。

隐藏是应对隐私危机的一种思路。目前加密工具盛行，通常以加密解决隐私泄露问题，但是加密工具也可能成为一种诱捕工具，可以以隐藏技术作为加密技术的补充解决这个问题。对于图片分享、外包、元数据挖掘等可能带来的泄露个人隐私的危机，都可以

通过隐藏通信,隐藏存储,隐藏社交、分享平台等手段来解决。



## 翁健： 大数据是解决其自身 带来的安全问题 的重要手段

暨南大学信息学院教授、院长翁健在论坛上作了题为“大数据环境下密码学新技术”的演讲。翁健教授认为,对于大数据的安全目标应从机密性、完整性、可用性3个方面来考虑。机密性指保证所有的敏感数据的安全性,即敌手无法从数据中得到除了长度之外的其他信息;完整性指保证未被授权的数据修改可以被检测出来,即恶意敌手不能在不被察觉的情况下修

改数据;可用性指敌手不能影响合法用户正常的访问数据和计算资源,通常使用密码学方法保护,如数据备份。

翁健教授指出,大数据带来新的契机的同时也带来了新的安全问题,但它自身也是解决问题的重要手段。我们需要以技术+政策法规的方式更好地解决大数据安全与隐私保护问题。

## 主题论坛：大数据学科与人才

5月25日下午，“大数据学科与人才”主题论坛在贵阳金阳万丽酒店5F黄果树厅举行，清华大学计算机科学与技术系教授，党委书记，教育部在线教育研究中心副主任孙茂松作为论坛主席主持了会议。孙茂松教授表示，大数据要能够真正地落地需要很多关键因素的促成，人才是其中关键的一环。没有成熟的学科体系和人才培养机制，就无法提供优秀的后备人才，因此以此为主题邀请了国内外专家学者从不同角度介绍和探讨大数据学科建设和人才培养的工作。会议现场气氛热烈，与会嘉宾与专业观众还进行了深入的互动交流。



主持人孙茂松教授与演讲嘉宾陈振冲教授交流



会场观众



演讲嘉宾与专业观众互动



## 孙茂松： 大数据人才培养机制 尚未真正形成

对于大数据人才培养上面临的挑战，孙茂松教授表示，当前大数据领域的人才培养机制仅是在现有框架体系下做了延展，并没有完全按照大数据包含的领域来设计体系，这也与大数据的领域特征有关。

首先，大数据领域比较广泛，多样性比较高，包括机器学习、并行计算、人工智能、统计学等多领域，同时还包括与工业领域的诸多结合，这样丰富的内容让大数据的课程设置很难全面覆盖，大数据的研究是相当庞杂的。

其次，大数据业务跨度不断扩大，涉及深度学习、基础学习、脑部深度计算、工程技术等，动辄都是几十亿的数据，且纵向延伸非常深，其研究需要极广泛的积累。

再次，大数据还具备很高的前沿性。基础科学、技术科学都是最前沿领域，采用最先进的技术，同时需要跟工业领域深度结合。如果重新设计大数据的课程体系，需要



孙茂松教授接受采访

选择上游某一段进行若干种组合，目前系统性的尝试仍未开始。



## 陈振冲： 数据科学家与数据分析师 是大数据人才培养的 两个重点

香港理工大学教授、学务长陈振冲教授认为，社会对数据科学人才的需求量非常大，目前学校提供的数据科学人才数量远不够。McKinsey Global Institute针对大数据人才需求所做的一个研究结果表明，到2018年，美国面临14万~19万的数据分析师缺口，150万的管理和运营人才。与数据紧密相关的产业都急需大数据人才，包括互联网、

电子商务、金融、工业服务、教育、娱乐文化、游戏、信息安全、O2O、社交媒体等。通过他们对中美两国对数据学科人才的招聘需求分析得出，目前市场所需的数据学科人才主要分为两类：数据科学家和数据分析师。对于数据分析师的学历要求相对数据科学家有所降低，而工作经验要求则高于数据科学家。综合来看，合格的数据数据分析师

需要具备较强的应用能力,收集管理数据;利用工具或软件分析数据并生成报告,撰写报告;利用工具或软件进行数据挖掘;实现不同的算法。合格的数据科学家需要具备

分析、研究、解决问题的能力;建立数据模型,分析模型;设计和实现数据分析、数据挖掘的算法;与商业或决策部门合作,利用数据分析结果提供决策支持。



## 翟成祥： 大数据课程设置要围绕 大数据应用

美国伊利诺伊大学香槟分校教授翟成祥认为,从教育理论来讲,大数据的价值将来很多方面体现在人的感知能力和判断能力,是一种扩充,是智能助手。大数据的课程设置应该紧密结合大数据的应用来设计,即4个A:第一个是acquisition(数据的获取),第二是aggregation(数据的整合),第三是analysis(数据的分析),第四个是application(数据的应用)。这4个A是一个循环。acquisition,涉及传感器的分布,如何优化获得的数据等;aggregation,涉及数

据库、搜索引擎、信息检索包括对数据的理解等;analysis,涉及数据挖掘和统计模型、高性能分析;对于application,目前比较可行的是用云计算的方式来支撑,因为数据比较大,另外一个发展方向是人工智能。

翟成祥教授表示,大数据时代已经到来,带来了挑战和机遇。挑战就是我们要培养什么样的人才,机遇就是可以改进教育,人才教育与教育方式二者结合起来,既能解决挑战,也能抓住机遇。



## 武永卫： 数据科学与工程“混合式 专业教育”

清华大学计算机系教授、副系主任武永卫介绍,2015年CSDN社区经过统计认为,未来3~5年中国所需的数据科学人才是180万,但是目前国内只有30万此类人才,缺口达到了150万。这说明了国内数据科学人才培养的紧迫性。目前大数据人才分为应用数据分析师、数据系统构架师、算法设计工程师3类。应用数据分析师具备行业应用领域知识,可提供基于数据挖掘分析的战略决策。数据系统构架师对数据的存储、分析与处理等技术和系统架构有深入理解。算法设计工程师则注重数据分析算法设计,统计分析

为核心的数学能力突出。

武永卫教授表示,针对大数据人才培养,清华大学设立了“数据科学与工程”混合式专业学位教育项目,其目的就是:培养具备数据存储、运行监管、智能分析挖掘以及战略决策等技能的专门人才,使其能够胜任数据存储管理师、数据分析师、数据系统构架师乃至数据科学家、首席数据官、商务分析师、战略管理者等职位。项目注重知识能力与职业素养融合的培养体系,设立了数据存储、系统架构、分析挖掘、应用/素养的相关课程。



## 陈学文： 数据科学人才要掌握 基本的技能

韦恩州立大学计算机科学系教授，大数据和商务分析中心主任，IEEE计算机学会计算生命科学技术委员会主席陈学文在论坛上做了精彩报告。

陈学文教授认为，数据科学的一个重要价值就是给商业带来价值，通过大数据挖掘、分析可以得到其价值。因此数据科学家要掌握的基本技能包括编程能力、数学思维、机器学习和可视化、强烈的好奇心、软技能（domain

knowledge and soft skill）。

陈学文教授介绍，目前美国在数据科学上开设课程的研究机构大大小小的有30多家，高校开设的课程都是针对硕士研究生的培养，如伯克利大学、纽约大学等。他建议应该有一个本科生的培养方案，这样学生在校期间跟着导师做项目，经过学习，学生会很熟悉整个流程，对学生的帮助很大。



## 熊贇： 数据人才培养应该向 精英人才培养模式转变

复旦大学计算机科学技术学院教授，上海市数据科学重点实验室主任助理熊贇在论坛做了精彩报告。

熊贇教授认为，数据科学家包括3类人。第一类人更多的是从事科学数据分析，第二类人是从事商业业务分析；第三类人是科研高校里的师生，即专门研究数据的人。

熊贇教授表示，当前国内很多学校都建立了大数据专业，但是整个学科体系还没有能够全面地建立起来。在培养过程中，领域专家重点是学习如何将领域业务需求转化为数据问题交给数据分析人员，并不关注数据处理细节；

而数据分析人员注重对领域专家所给的数据集进行处理，缺乏对领域知识的理解。这是目前数据科学家培养的常见方式，但却缺乏系统性。需要一个新的学科体系，包括数据科学基础理论、科学研究、数据方法、数据界探索、数据技术及应用、领域数据学。

熊贇教授表示，数据科学并不是简单的学科交叉，应该基于并和所有学科相关。而超学科数据人才培养体系，应该实现以团队培养为主的数据人才培养模式向培养具有数据能力的人（而非团队）为目标的培养模式转变。

（内容整合了嘉宾演讲速记及相关媒体报道）

大会网址：[bdse2016.j-bigdataresearch.com.cn](http://bdse2016.j-bigdataresearch.com.cn)

## 收录声明

凡投本刊并被录用的稿件文章,若无特别声明,视同将录用文章的电子网络版、汇编或合订本的使用权授予本期刊社,本刊支付的稿费中已包含上述费用。

If the authors of the accepted article by Big Data Research do not provide special announcement, the editorial department believes that the authors will be granted the right of their article to Big Data Research. The expenses have been included in the fee paid by editorial department.

## 道德声明

本刊发表的论文是作者独立取得的原创性研究成果,无一稿多投;论文内容不涉及国家机密;未曾以任何形式用任何文种在国内外公开发表过;论文内容不侵犯他人著作权和其他权利。若发生一稿多投、侵权、泄密等问题,论文作者将承担全部责任。

The authors of Big Data Research guarantee that their submitted article is original and contains nothing confidential. The said article is only submitted to Big Data Research. The said article has not been published before and has not been submitted elsewhere for print or electronic publication consideration. The said article is in no way whatever a violation or an infringement of any existing copyright or license from the third party. Otherwise, the authors of the said article shall take the blame for the violation or infringement of the related copyright and the leakage of secrets.

## BIG DATA RESEARCH

(Bimonthly, started in 2015)

Vol.2 No.4, July 2016

**Competent Unit:** Ministry of Industry and Information Technology  
of the People's Republic of China

**Sponsor:** Posts & Telecom Press

**Academic Guidance:** CCF Task Force on Big Data

**Publisher:** China InfoCom Media Group

**Address:** F8, You Dian Publisher Building, No.11 Chengshousi  
Road, Fengtai District, Beijing 100078, China

**Manager:** Liu Hualu

**Deputy Manager:** Yi Dongshan, Liang Haibin

**Chief Editor:** Zheng Weimin

**Director:** Wu Nada

### Editorial Committee

**Adviser:** Li Guojie, Wu Hequan, Li Deyi, Fang Binxing, Mei Hong

**Chairman:** Zheng Weimin

**Vice Chairman:** Cheng Xueqi, Du Xiaoyong, Sun Maosong,  
Liu Duo, Wang Xiaodan

**Tel:** +8610-81055448, +8610-81055475, +8610-81055490

**Fax:** +8610-81055494

<http://www.j-bigdataresearch.com.cn>

**ISSN** 2096-0271

**Date of Publication:** 2016-07-20

**Price:** ¥ 35.00



# “大数据标准”征稿通知

“大数据”不仅仅是一代新的信息技术的代名词，也是解决复杂问题的一种新方法。大数据的发展和应用已经对政府、企业决策和人们的生活方式产生深远的影响。目前，各种大数据技术和系统层出不穷，大数据应用方兴未艾。然而，大数据系统之间缺乏有效的互操作手段，大数据的开放共享受制于安全与隐私方面的顾虑，大数据推广应用缺乏工程化的指南，在很大程度上阻碍了大数据的发展。甚至大数据的定义、相关术语、分类、架构等方面也还缺乏统一描述，各种大数据产品的技术要求也不尽相同。标准化无疑是解决上述困境的有效手段，然而，无论是国际还是国内，迄今为止其标准规范的研究工作都还很匮乏。因此，加强大数据标准化研制，对推动我国大数据产业进程，加快技术与标准的相互融合，推动大数据发展国家战略和数据立法具有重要意义。

本刊将以“大数据标准”为主题进行征文，以凸显国家及行业对大数据标准的重视性及聚焦点，发表国内外大数据标准在建立和执行过程中具有典型性和创新性的高水平研究成果，深入探讨大数据标准的未来构建方向。

## 一、征稿范围

征稿范围包括但不限于：

1. 大数据标准体系构建的理论方法、框架和系统；
2. 大数据标准的国际前沿动向分析与展望；
3. 大数据系统的技术标准研究；
4. 大数据标准的行业建设案例介绍；
5. 大数据安全与隐私保护标准建设；
6. 大数据开放共享的机制与标准研究；
7. 大数据的定义、相关术语、分类、架构等方面的标准化研制；
8. 大数据技术标准与基准测试平台。

## 二、征稿要求

1、文稿应属于作者的原创性科研成果，数据真实可靠，具有重要的学术价值与推广应用价值；应未在国内外公开发行的刊物或会议上发表或宣读过，也不在其他刊物或会议的审稿过程中，不存在一稿多投问题。

2、文稿应包括中英文题名、作者信息、摘要、关键词、基金项目名、参考文献，中文正文和作者简介（包括照片），文稿一律采用 word 文档提交。

3、投稿时请注明作者联系方式（电子邮件、电话、邮寄地址）及“‘大数据标准’选题投稿”及所属方向。

4、文稿通过《大数据》网站（<http://www.j-bigdataresearch.com.cn>）进行投稿，暂不支持其他投稿方式，投稿须知及文稿模板请参见网站说明。

## 三、专题特邀客座编辑

杜小勇 中国人民大学信息学院教授

吴东亚 工业与信息化部电子工业标准化研究院信息技术研究中心副主任

## 四、重要日期

文稿截稿日期：2016年10月20日

计划出版日期：2017年3月

# IMAGE TARGET DETECTION AND APPLICATION

## 图像目标检测

## 技术及应用



本书系统地阐述图像目标检测的有关概念、原理和方法。在内容上既选择了有代表性的图像目标检测的经典内容，又结合复杂场景中所面临的困难和问题及作者近年来关于图像目标检测关键技术的研究与应用实践，选取了一些新的研究成果，具有一定的针对性、广度、深度和新颖性。

本书内容新颖，理论联系实际，可作为大专院校及科研院所图像处理、计算机视觉和视频处理等领域的高年级本科生、研究生的教学和参考用书，也可供相关领域的教师、科研人员及工程技术人员作参考。

出版社：人民邮电出版社

书号：978-7-115-41879-1

联系电话：010-81055480

# 数聚新动能 数创大未来

BIG DATA - NEW MOMENTUM, GREAT FUTURE

## 2016 中国国际大数据大会

2016年9月27-28日 北京·富力万丽酒店

全景式交流展现大数据前沿热点



政策



市场



标准



技术



营销



应用



交易



安全

2

场

峰会

12

场

论坛

100

场

演讲

2000

人

规模

主办单位：人民邮电出版社

协办单位：中国互联网协会推广与普及工作委员会  
信息通信大数据产业联盟

承办单位：信通传媒、数创汇

支持单位：《大数据》杂志

☎ 010-62882294 www.cbds.com.cn



官方公众号



报名参会



# 传递大数据产业价值思想与趋势

## 中国国际大数据大会开幕式暨全体大会

- 01 大数据与网络安全论坛
  - 02 运营商大数据论坛
  - 03 中国大数据技术方案商大会
  - 04 大数据助力政府治理与智慧城市发展论坛
  - 05 工业大数据与中国制造2025论坛
  - 06 数享家峰会（主会场）
  - 07 大数据与数据中心发展论坛
  - 08 大数据交易高峰论坛
  - 09 大数据与分享经济
  - 10 互联网+大数据峰会
  - 11 大数据金融与征信论坛
  - 12 大数据与旅游业创新发展论坛
  - 13 大数据双创路演
  - 14 品牌发布会
  - 15 产业报告发布会
- 大数据创新展 大数据图书展**

# DPDK 应用基础

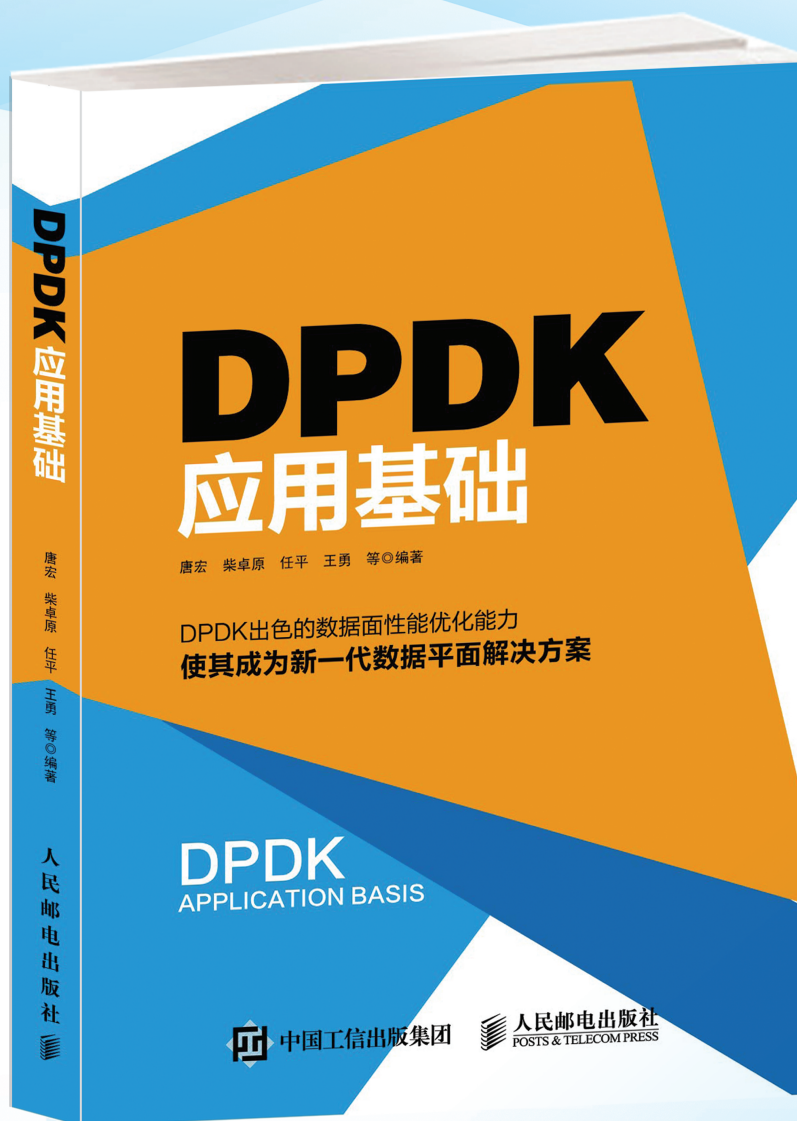
唐宏 柴卓原 任平 王勇 等◎编著

DPDK出色的数据面性能优化能力  
使其成为新一代数据平面解决方案

人民邮电出版社出版

书号：978-7-115-42604-8

出版日期：2016年8月



DPDK 不仅是一个用户态的 I/O 加速方案，也包含了 NFV 开发测试环境中所需的高速网络应用样例库以及开发调试和测试工具集。目前有越来越多的开源项目或商用解决方案基于 DPDK 进行了实现，DPDK 已经成为 NFV 加速的基础性架构和公认的应用加速接口。

本书内容分为基础原理、DPDK 应用与测试和应用开发及实例解析三个部分。其中，第一部分简要介绍了 DPDK 相关背景、技术原理、部分库函数、DPDK 安装部署和调试方法，帮助读者从宏观上了解 DPDK，对其有一个基本认识；第二部分阐述了 DPDK 技术在 NFV 场景下的应用，并通过 16 个测试用例说明了基于 DPDK 的 NFV 转发性能调优实践与测试结果，有助于业界同仁开展 NFV 系统的研发与性能评估工作；第三部分介绍了基于 DPDK 的应用案例，案例介绍中融合了对如何基于 DPDK 进行上层应用开发的解析，以帮助读者更加深入地理解 DPDK。

本书可以作为网络技术人员和 IT 系统开发人员的中初级读物，帮助他们快速了解 DPDK 社区及其大型通用控制器的开发方法。

# 科学数据 与创新发

## 2016第三届科学数据大会

8月25日-26日，上海

研讨科学数据与创新发，关注科学数据在学科创新、管理创新和产业创新方面的作用、途径、实践和挑战。

主办：国际科技数据委员会（CODATA）中国全国委员会

共同主办：国家科技基础条件平台中心 中国科学院条件保障与财务局

承办：上海市数据科学重点实验室 《大数据》期刊

中国科学院计算机网络信息中心

会议网址：<http://dc2016.codata.cn>



邮发代号：2-537 国外代号：C9118 定价：35.00元

ISSN 2096-0271



9 772096 027162

