



# 数据分析

CHINA DATA ANALYSIS 用数据说话·做理性决策

++ 中国商业联合会数据分析专业委员会 主办 ++

中国数据分析会员特刊  
2015年第03期 总第23期  
咨询热线：010-59000991  
[www.chinacpda.org](http://www.chinacpda.org)



关注行业微博微信 了解数据分析行业前沿知识

## 小心! WARNING! 你不是别人的鱼!

协会带你揭示某机构不可告人的秘密



p14. 行业热点 - 协会带你打“假CDA”

p06. 行业动态 - 第三届数据分析行业峰会回顾

p22. 新兴项目数据分析师事务所如何走好创业之路

p24. 2015年大数据展望和市场预测

p33. COX回归模型在电信用户离网分析中的应用



《中国数据分析》会员特刊  
2015年第03期 总第23期

### 主办

中国商业联合会数据分析专业委员会

### 主编

张楠

### 编委

冯伟 李子昂

### 出版时间

2015年10月

### 美工

崔峻珩

### 联系我们

中国商业联合会数据分析专业委员会

地址: 北京市朝阳区朝外soho C座9层 100020

电话: +86-10-59000991

传真: +86-10-59000991转 607

### 投稿

欢迎广大读者踊跃投稿, 内容包括学术观点、教学体验、教学活动、学习感悟、实战经验、随笔文章等。稿件附图格式为JPG或TIFF格式, 大于1M, 分辨率在300dpi以上。

感谢您对《中国数据分析》的支持!

投稿邮箱: marketing@chinacpda.org

## 目录 CONTENTS

### P01 卷首语

硬道理+软实力=?

### P02 行业动态

洞悉行业大事, 量察热点潮流!

一、数据分析公益沙龙活动全国各地精彩召开

二、河北省第一期项目数据分析师课程开班仪式在石家庄举行

三、新面孔、老朋友 —— CPDA项目数据分析师、CDA数据  
分析员官网更新升级

[峰会回顾] 触摸大数据本质 助力企业数据化发展

[峰会主题演讲精选] 触摸大数据本质

[峰会主题演讲精选] 大数据人才对零售业未来的影响

### P14 行业热点

关于"CDA注册数据分析师协会"在华运营的调查  
协会为CDA数据分析师正名

经济日报: 想考CDA数据分析师, 需擦亮眼睛  
大数据时代, 技术和分析哪个更重要?

为什么CPDA项目数据分析师不分级?

国务院印发《促进大数据发展行动纲要》

工信部正制定《大数据产业"十三五"发展规划》

### P22 会客厅

新兴项目数据分析师事务所如何走好创业之路

### P24 "数"业专攻

2015年大数据展望和市场预测

后Hadoop时代的大数据架构

### P32 运"数"有道

零售企业库存管理分析

COX回归模型在电信用户离网分析中的应用

### P35 事务所风采

河南智宸项目数据分析师事务所

山东海洋项目数据分析师事务所





**CDAC**

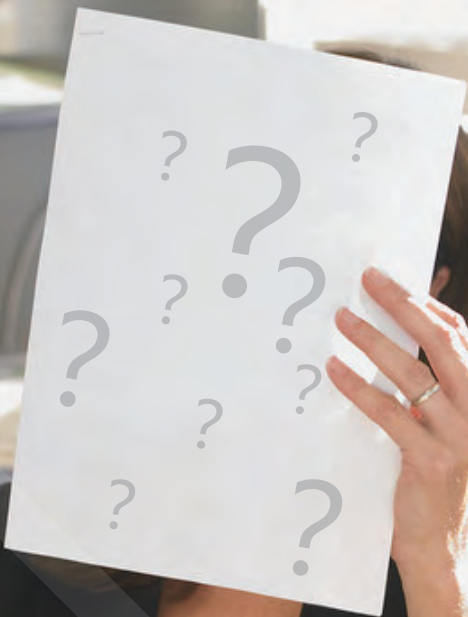
大数据的海洋里，你还在孤舟漂泊么？  
感到迷茫？无助？不知如何才能更好的前行？  
快来成为协会的一员吧，您将找到寻求的一切！  
我们携手一起  
**乘风破浪，扬帆起航！**

咨询热线：010-59000991转651（会员处）

中国商业联合会数据分析专业委员会为中华人民共和国民政部正式批准和登记的中国数据分析行业唯一的行业协会。



CPDA



漫无目的  
不如  
**精准定位**

项目数据分析师  
引领大数据时代的  
知识体系变革

早培训

早升职

早加薪



给自己一次机会，  
迎接无限精彩的未来！

提升量化经营技能，掌握数据分析必备的科学方法论！  
正确理解、灵活运用于企业决策全面数据化分析中！  
成功融入大数据时代，释放数据分析师全局视野！  
行业协会权威认证，承载坚实的启航平台！



行业协会  
官方微博



官方微信：  
wxchinacpda

★ 关注微博微信参与互动 赢好礼！



[www.chinacpda.com](http://www.chinacpda.com)



400-050-6600

## 硬道理 + 软实力 = ?

任何事物的发展壮大都要遵循一定的客观规律，有因必有果，因为我们需要它的存在和发展，它才会不断的变化和升华，因为我们需要了解它的本质和表象，它才会不断的被研究和发现。大数据就是那个被我们不断追求和关注的它。

硬道理决定了大数据时代的到来，也注定了这个行业的发展。2014年“大数据”首次出现在《政府工作报告》，其指出：要设立新兴产业创业创新平台，在大数据等方面赶超先进，引领未来产业发展。“大数据”随即成为当年两会的热议词汇。在今年8月19日的国务院常务会议上，通过了《关于促进大数据发展的行动纲要》，李克强总理用三个关键词“共享、开放、安全”再次提到了对大数据的展望。


在中国，政府的决心和态度一向都是绝对的硬道理，其实一些西方发达国家很早就纷纷制定了大数据战略。例如，在2013年美国奥巴马政府宣布投资2亿美元拉动大数据相关产业发展，将“大数据战略”上升为国家意志，并且将数据定义为“未来的新石油”。由此可见政府对于大数据产业的期望和憧憬。除了政策导向，另外一个硬道理就是“刚需”，有需必有求。在中国这个拥有10多亿人口的国家，每天产生的数据量是无法估算的，大到我们无法用语言和数字来形容和展现，每个企业每秒都会产生数据，大数据的采集和分析对于任何一家企业的决策发展都是关键的，现在已经有越来越多的企业意识到了这一点，例如李彦宏、马云、雷军等IT业界大佬，就是大数据概念的拥趸，并已将大数据的构想付诸行动。在大环境的影响下，一些传统企业也纷纷加入了“算数”的行列。

有了硬道理作为坚实的发展基础，我们要做就是不断加强软实力。所谓软实力，包括：数据分析技术的更新，数据分析人才的培养，数据分析机构的建立，还有就是数据分析行业的规范。R语言，hadoop等各种算法应用不断被开发更新。在10多年的时间里，协会已经培养了近万名项目数据分析师，考察审批了上百家项目数据分析师事务所，但是这些数字还远远达不到行业发展的需求。与大数据概念知名度和企业热情形成对比的是：大数据正面临全球性的人才荒。企业对新型大数据分析和预测技术人才的热情和需求正在超过传统的商业智能和信息管理人才，就业前景一片大好。面对一个如此朝阳的行业，作为行业协会，我们鼓励更多的人员成为项目数据分析师；鼓励更多的项目数据分析师创业，成立事务所；鼓励所有对数据分析感兴趣的人加入这个奇妙而未知的领域。同时我们也要秉承可持续发展的原则，在百花齐放的同时，共同遵循行业规范，引领整个行业能够健康有序的发展，真正的能够为大家带来帮助。软实力就好比一种修炼，需要我们潜心研究，不断积累。

在我看来：硬道理和软实力相辅相成，缺一不可。

硬道理+软实力=大数据时代的到来。

中国商业联合会数据分析专业委员会







## 洞悉行业大事，量察热点潮流！

文 / 中商联数据分析专业委员会市场处 张楠 图 / 崔峻珩

### 一、数据分析公益沙龙活动全国各地精彩召开

北京：大数据的“行业误区”与“职业规划”



2015年8月29日，由中国商业联合会数据分析专业委员会举办的数据分析公益沙龙完美落幕。六十余名来自不同行业的数据分析师和数据分析爱好者齐坐一堂，参与了这次数据分析领域的饕餮盛宴。

这期沙龙在北京CBD核心地带—财富中心千禧公寓咖啡吧举办，特邀了中国商业联合会数据分析专业委员会会长邹东生老师；资深数据分析师和软件架构师、中国商业联合会数据分析专业委员会数据中心主任何林老师为主讲。基于数据分析的深度技术与行业知识，向大家解惑“大数据”的行业误区与职业规划。

邹东生老师以“数据分析——把握大数据时代的风口”为主题，通过对数据分析人才、数据分析技能、咨询与技术等多方面深入剖析并清晰阐述了数据分析行业目前三大问题的解决方案——数据分析人才对企业至关重要；真正的数据分析技能，重心在深入与分析；以及数据分析可以通过技术加咨询的方式创造价值。

何林老师以“数据分析师成长之路”为题，毫无保留的为大家传授了他作为资深数据分析师的职业经验、发展道路、优势和机遇。并说明了统计学知识、R语言与python在数据分析中的实际应用方法。在细致的讲解之后，大家纷纷表示听的非常过瘾。于是何老师又兴致勃勃的现场加讲，结合实际技术与商业案例，为大家具体阐述了解决实际问题的数据分析思路

与方法。

经过四个小时的精彩讲座，这次沙龙在诸多朋友们的热烈讨论中落下帷幕。会后，参与本次沙龙的数据分析爱好者们纷纷表示此行收获颇丰，意犹未尽。而我们接下来还会有一系列干货十足的公益沙龙活动，欢迎广大数据分析爱好者共同加入，与我们一起推动数据分析行业的蓬勃发展！**FIN**

## 河南：数据分析与工作

2015年9月20日下午14时，郑州的数据分析师们欢快地迎来了：“数据分析师沙龙”活动。

本次活动是项目数据分析师河南授权中心主办，河南明豫项目数据分析师事务所承办，河南龙米科技协办的一次“数据分析”盛宴。

这次活动是河南明豫项目数据分析师事务所的大数据时代讲堂的一部分，在郑州市郑东新区绿地之窗云峰座B座28楼大会议室举行，本次活动得到天机创投机构提供的会场以及设施支持，也得到了河南CPDA学友会的支持，河南明豫项目数据分析师事务所和龙米科技为本次活动提供了主要礼品和奖品的赞助。

### 有干货



下午14时许，主持人欢迎大家的到来。随后，学友会的周彦峰同学进行关于《服装行业的销量预测》的报告，通过河南某服装销售企业的案例，做了一次贴近实际业务的案例分析和梳理，以预测销售量为主要思路，给大家分享讲解了灰色系统的概念和应用、多次线性回归的选择和实际业务的结合等知识，加深了大家对知识点的掌握和运用。其中很多同学很关心多次线性回归的选择问题，并进行了深入的探讨。

### 有红包

在周彦峰同学演讲问答过后，由河南明豫项目数据分析师事务所的孙春光，给大家现场建立了微信群，发放了人手一

个的微信红包，把整场活动推向了第一次高潮。

### 有冲击力



下午15时，河南明豫项目数据分析师事务所的孙春光给大家做了《大数据时代商业模式的应用》的主题分享，通过讲解大数据时代分析体系的主要特点、数据积累驱动企业优化变革、拓展新业务发展领域的雅昌艺术网成功案例、数据分析建模锁定客户，改变Target等零售巨头的业务增长模式的成功案例、河南明豫客户的市场细分成功案例等一系列案例的讲解，突出数据分析人才和数据分析技术，以及数据分析专业机构在企业变革发展道路上举足轻重的地位，给到来的数十位嘉宾以及学友、企业人士，很强的冲击力。在结束后的互动环节，大家畅想企业在发展过程中数据分析人才和能力的重要性，现场气氛进一步推高。

### 有红包不断档，礼品节节高！

在孙春光的主题演讲之后，大家又开始抢红包活动，这次抢红包把沙龙的气氛彻底推向最高峰。

### 有需求



下午16时，河南龙米科技总经理孙继龙先生作为此次活动的特邀嘉宾，开始了《数据分析在微信营销中的应用》的主题分享，他先用数据介绍了社会大众对微信功能的依赖和在营销中的巨大威力，然后对微信营销领域数据分析的尝试和成功案例做了介绍，在结合自己业务的微信营销实务里，他谈到：“现在我们的数据量越来越大，一般的数据展示性分析已经不



能让我们在这一领域里有所优势了，欢迎在座的各位数据分析师们与我们合作，开展深度分析服务的业务。在座的数据分析学员们看到数据分析的春天已经到来了。大数据和数据分析就在我们身边。

### 有礼品，更有内涵！

演讲勾起了大家对创新高效营销方式的探讨，孙继龙先生也给在座的听众准备了很好的礼物，让大家忘记了疲劳，兴趣高涨的继续我们的沙龙活动。FIN

## 上海：数据分析在工作中的应用场景

2015年9月19日，上海CPDA授权中心的数据分析沙龙活动成功举办，此次沙龙活动的主题为“数据分析在工作中的应用场景”，此次沙龙邀请了有丰富数据分析实战经验，正在进行大数据创业的胡子欣博士进行分享。上海CPDA学员和广大数据分析爱好者因为喜欢数据又聚在了一起。



胡子欣是复旦大学生物信息学博士。主要研究方向图像识别、泛函分析、分类器设置与优化。曾在尼尔森工作，主要侧重消费数据的数据分析及咨询服务。胡子欣博士以他非常丰

富的企业数据咨询案例经验，总结出来一系列的如何做好数据分析真实体会，数据分析是一个有标准逻辑以后的精细化过程，主要过程就是问题数据化，然后解析。

他用三个精准化营销的案例给大家阐述了，数据分析一定要有明确的目标，明确的计算逻辑并有明确的解答。在数据做KPI考核时，做客户定位非常重要，即根据数据了解什么产品给什么客户，什么客户应给他推什么信息，如何通过人物画像城市画像等分析具体的事物和人才对。

同时，他也同参加沙龙的同学交流了自己对大数据发展方向的个人看法，并给大家分享目前自己创业所重点提出的全景CRM观点，即由CRM体系迭代带来的aggregation层次更新，对很多在场人员很多启发。

嘉宾老师分享后，参会同学们都根据自己所在行业中遇到的数据分析问题纷纷加入讨论，并与嘉宾老师互动、探讨。参会同学之间也互相认识、交流、讨论。上海CPDA通过这种线下的沙龙，可以让更多企业认识到数据分析的价值并促进不同行业的数据分析师们信息和经验分享。FIN

## 二、河北省第一期项目数据分析师课程开班仪式在石家庄举行



2015年7月25日，中国项目数据分析师(英文全称CERTIFIED PROJECT DATA ANALYST，简称CPDA)河北授权管理中心授牌暨河北省第一期项目数据分析师课程开班培训仪式在石家庄举行。中国商业联合会数据分析专业委员会秘书长蒋文伟为河北授权中心颁发授权证书与铜牌。据悉，现在国内有100多家项目数据分析师事务所，北京有20多家，河北却一家都没有。河北项目数据分析师授权管理中心的成立，填补了该省没有项目数据分析师的空白，开启了河北“大数据”时代的新篇章。





该中心负责人成丽珊介绍，项目数据分析师(CPDA)是专业从事投资和运营项目数据分析的高级决策人，他们能够通过掌握大量行业数据以及科学的计算工具，将经济学原理用数学模型表示，科学合理的分析投资和运营项目未来的收益及风险情况，为做出科学合理的决策提供依据。而研究指出：50%的总经理、营销总监都无法获得及时、完整的销售经营数据、信息，更有90%的人不了解、或是根本不会使用已经呈现在他们面前的销售数据。随着企业快速成长扩大，产品线增加，营销人员增多，区域分散，业务复杂，在销售数据、信息的整合上变得开始混乱、无序起来。

今天，在大数据时代、互联网时代，仅仅依靠一些常规的营销分析报表，已经不能满足企业管理者的决策需求，于是项目数据分析师(CPDA)应运而生。据了解，报名参加第一期项目数据分析师(CPDA)的学员大多来自河北、山西两地，从事行业涉及金融、物流、通信、保险等，参加培训的人大多是企业高管、单位财务管理、市场决策、企业管理人员。参加课程的学员都在热议“大数据时代来了，相信在未来的三到五年，数据分析行业必定持续火爆呀”，我们现在参加学习数据分析将会给我们在工作中得到很大帮助。第一期项目数据分析师(CPDA)河北地区的成功举行，将是河北地区自主培养项目数据分析人才的开端，是河北地区数据分析发展的重要一步，这必将带动整个石家庄，整个河北省数据分析工作的快速发展，为京津冀大数据的发展提供优秀人才。FIN

### 三、新面孔 老朋友——CPDA项目数据分析师、CDA数据分析师官网更新升级

为了更好的提升服务质量，增强学员体验，及时了解数据分析行业、学习动态，7月份协会对 CPDA和CDA两个官方

网站都进行了优化升级：选用了更为先进和快速的服务器，响应速度大幅度提升；页面相较于之前排版更为人性化、内容更加丰富全面、风格更为大气美观。

新版官网页面更为简洁，行业资讯、协会动态尽收眼底；内容更为丰富，更加注重学员感受，将开课、培训、学习导航等的最新动态信息放在页面主体位置，为了增强学员们的使用体验，及时得到学员们的反馈信息，在页面主体的右侧设置了互动专区，以便随时与学员互动交流；除了信息发布实时化，内容更新需求化。学员还可以全方法的了解CPDA和CDA课程的各种信息，包括，资质、证书、常见问题、数据技术，招聘信息、学员风采和学习心得等。



虽然我们对以上两个官网都进行了升级和改版，但随着协会人才培养计划逐步完善、服务质量不断提高，我们的官网仍将不断改进，同时也欢迎广大学员随时与协会、培训处、继续教育处进行互动，提出宝贵建议，我们将针对学员需求不断优化官网质量。FIN



## [峰会回顾] 触摸大数据本质 助力企业数据化发展

编辑 / 中商联数据分析专业委员会市场处 张楠 图 / 崔峻珩

第三届中国数据分析行业峰会已于2015年7月10日在江苏昆山落下帷幕，本次峰会除了现场精彩的举办盛况，还有许多难忘的瞬间，以及许多值得回味的内容。就让我们一起回顾这次行业盛会，一起触摸大数据本质！

被业界誉为“知识和思想盛宴”的中国数据分析行业峰会，自2010年开始，至今已成功举办两届，得到社会各界的重视和广泛关注。

今年7月10日，第三届中国数据分析行业峰会在江苏昆山举行，这是一次数据行业的盛会和数据分析人才的高端聚会。作为第十届中国零售商大会的分论坛，有超过1600名来自零售、IT、教育、金融等行业的代表参与本次峰会。会议围绕“触摸大数据本质”，探讨大数据本质及数据人才、技术应用等前沿话题，分享大数据应用“落地”案例，为各行业在数据平台的实际应用中遇到的问题寻求解决途径。

大数据概念从2003年开始正式提出来以来，在各个领域影响和改变着我们的生活。10号的峰会论坛上，来自行业主管协会、数据分析专业事务所、大数据技术企业等12位业界专家和数据分析科学家，与大家分享了关于技术、人才、量化决策等主题演讲。

在上午的主题论坛中，嘉宾演讲主题主要集中在“大数据本质”、“大数据如何运用”和“数据分析人才”领域。下午的嘉宾演讲，则以关注行业案例为主，演讲嘉宾是来自高端乘用车行业、社交媒体研究、互联网金融等领域的大数据研究者，分享了他们在各自专业领域中数据应用的实际案例。





峰会现场，来自中国商业联合会数据分析专业委员会、百度研究院大数据实验室、北京大学商务智能中心、台湾云图科学计算股份有限公司的研究人员、数据分析科学家为现场200多位听众献上了12场精彩的主题演讲。

际问题，都可以通过微信群提出讨论。“高峰论坛”进行时，还会在大屏幕上直播现场观众的问题。

2015（第三届）中国数据分析行业峰会正式开始！主持人正在介绍今天会发表演讲的重要嘉宾和会议议程，会场里座无虚席！



大会签到处，参与本次峰会论坛的嘉宾们开始在这里签到了。一场数据分析行业的高端人才聚会和行业大会即将拉开帷幕。

活动主办方中国商业联合会数据分析专业委员会会长邹东生先生发表开场致辞。邹会长用“很热”和“很浮躁”来形容目前中国大数据市场情况。



签到后加入微信群，可以参与互动、更加方便了解峰会进程。对大数据有什么看法，在企业的数据化进程中遇到的实



他举了一个例子：今年，新加坡政府通过协会搭桥，希望吸引一批优秀大数据人才到新加坡就业。现在宣传大数据、炒作大数据概念的人很多。但同时，在火热的背后，中国企业还非常缺乏应用大数据实实在在帮助企业成功的案例。总结大数据本质，邹东生会长提出：可以记录的是数据，能够给企业提供依据和帮助的才是分析。大数据在中国的真正落地，可谓任重道远。

协会数据中心主任何林在《大数据用起来》的主题演讲中指出：当前中国企业遇到最大的问题是如何让大数据落地，为其所用。没有建立数据化管理的企业面临如何解决数据有效

存储；有数据没有应用的企业，考虑如何打通内部数据、搭建分析平台；而正在进行数据应用的企业则考虑如何有效解决算法和难点，从而转化为商机。



来自台湾云图科学计算股份有限公司的创始人兼首席数据科学家周庭锐先生的演讲围绕“自动计算”这个话题。



“自动计算”是大数据应用里的一个不常被讨论，却隐匿在任何商业模式之下，几乎无所不在。小至无人飞行器的悬空拍照，大至 PM2.5的明日预报推算，处处隐含着“自动计算”的影子。周庭锐教授分别从“应用对象”、“计算方法”、与“数据结构”三个视角出发，由简至繁讨论三种不同层次的自动计算方法。

峰会重头戏之一，也是非常具有开放性的一个环节“高峰对话——数据分析对商业模式创新的革命性影响”举行。

莅临本场对话的嘉宾：中国商业联合会数据分析专业委员会会长邹东生先生，台湾云图科学计算股份有限公司创始人兼首席数据科学家周庭锐，银联智惠联合创始人兼CTO龙凯先生，中国商业联合会数据分析专业委员会数据中心主任、首席数据分析科学家何林先生，峰会主持人GE金融前亚太区董事总经理、天得一清投资管理有限公司总裁朱叶青先生。



上午的演讲继续进行，中国统计信息中心大数据研究实验室主任江青女士，她重点从企业竞争力的角度分析了发数据。她表示：无论是政府部门，还是商业部门，包括科学、医疗、广告等领域，都无法忽视这一由于技术进步所带来的变化。互联网加大数据使传统行业跨界成为了一道新的风景，近年来的实践表明，大数据成为了企业营销的致胜竞争力，企业营销离不开数据参考也已经成为一种常态。



项目数据分析师上海授权中心负责人徐晓颖的演讲题目是《数据人才需求与培养》：2014年最炙手可热的25项工作技能中，排名第一位的就是“统计分析和数据挖掘”能力。





但目前国内企业数据化程度低，有超过50%的企业还没有成立数据分析部门。企业、高校和社会组织等对于复合型的数据人才的需求量十分巨大的。一套完整、高效且实用的数据人才培养和认证系统，可以有效帮助有志加入数据分析行业的人员实现愿望。



接下来，中颢润（北京）项目数据分析师事务所副总经理王庆生先生分析了目前零售业面临电商的冲击导致规模萎缩，增长率下滑的现状。在大数据时代下，通过系统性的，体系化的数据分析方法，不仅能在策略制定的完备方面给出指导意义，同时，在运营期间也从产业链与供应链的综合分析，给出科学的决策。



作为我国数据领域唯一的行业组织，中国商业联合会数据分析专业委员会通过拥有自主知识产权的 Datahoop 数据分析平台，为企业免费提供技术支持、降低大数据应用的门槛。

第三届中国数据分析行业峰会现场，“2015-2020年百家企业经营模式创新与转型战略合作”项目启动，正式开展这个长达5年期的企业大数据公益扶持项目。企业正式签约后，能够获得免费扶持期6个月或1个完整项目周期。

“百家企业扶持计划”从2015年6月1日开始至2020年5月31日截止，不论是从海外上市公司，还是在小微企业，都能够申请参加。上午峰会结束后，现场嘉宾在展台领取“百家企业扶持计划”资料，详细了解计划和加入条件。

10号下午的演讲，以关注行业实际案例为主，来自百度研究院大数据实验室的数据科学家沈志勇先生，介绍了大数据智能分析技术在搜索引擎公司从网页搜索、自动问答、推荐系统、语音识别、预测系统等领域的全方位应用。



他还提出：目前从信息化到数据化，最终到智能化的过程中数据化程度低，以及如何保护数据隐私依然是阻力之一。

北京大学商务智能中心研究员黄丹阳博士，重点关注了大数据分析在社交媒体研究方面的应用。



随着互联网技术成熟，网络文本数海量增加，如何针对数据进行分析，并进行统计推断也成为了当前的研究热点。社交网络中用户的标签是用户自己的定义，并且反映了用户的爱好、职业、生活方式等特征。通过研究标签，可以建立了用户个性化标签和社交网络结构之间的模型，并能够应用于用户个性化标签推荐。



中国国际电子商务中心内贸信息中心副总经理李正波提出：随着云计算、大数据、物联网等技术的快速发展，数据分析与应用在经济、社会、政治、等领域的作用日益重要。利用大数据技术，深度挖掘数据之间的关联关系，进行全方位、智能化的分析计算和可视化的结果展示，可以为商务领域的政府、企业等单位开展决策分析等工作提供更加全面的数据支撑和技术保障。



北京大学商务智能中心研究员朱雪宁博士利用大数据分析办法，对100万在线用户的13亿搜索序列文本做了分析，并对高端车用户以及商学院人群做了对比分析。通过提取有效指标、数据分析以及统计模型的方式，分析高端车潜在用户搜索平台上表现出的行为，从而对在搜索引擎中的广告投放的策略提出可行建议。

银联智惠 CTO 龙凯在演讲中分享了数据价值三定律：

- 1、沉默数据急速贬值
- 2、流动产生价值
- 3、融合促进增值 $1+1>2$ ,

并且分享了这三条定律在商业智能、互联网金融与安全

云上的运用案例。



湖南翰林项目数据分析师事务所所长卿启伟认为在大数据扑面而来时，我们更需要一种理智的心态去看待大数据。那种认为跨入大数据行业就会大捞一把的想法是很天真的。



大数据时代不是比谁拥有的数据最多，而是比谁驾驭和运用这些数据并将其对接市场需求的能力和手段最多，需要经营者用心寻找，用心体会。FIN





## 【峰会主题演讲精选】触摸大数据本质

编辑 / 中商联数据分析专业委员会市场处 李子昂 图 / 崔峻珩

介绍：中国商业联合会数据分析专业委员会会长、北京大学光华管理学院MBA校友导师、北京大学光华管理学院MBA、北京市青联第十届委员。同时，也是中国数据分析行业发起人、奠基人，具有丰富的企业经营管理咨询经验，是资深数据分析专家，曾主持编写《投资数据分析》、《经营数据分析》等书

非常高兴跟大家一起，在这里老朋友也有新朋友，此次峰会已经第三届了，我们借助零售商大会和大家一起探讨行业的发展。今天在座的12位嘉宾的演讲都会很精彩，这次会议的主旨是“触摸大数据本质”，所以我们请到了行业的专家、实践者们，他们有很多能够让大家兴奋和感兴趣的话题。我今天简单跟大家交流几个小问题。

第一个是这两年大数据非常火热，有几个小事情可以看出。昨天中国零售商大会也跟大家交流过近期的国务院6月24号发布了大数据各种指导性意见的文章，里面提出很多支持大数据发展的意见和建议，政府给予了越来越多的支持。第二个是去年一个小事情但是能说明一些问题，就是我们接到了国外政府人力资源局、经济发展局的接洽，希望大力推动引入大数据人才到国外工作，甚至移民。通过此事大家可以看到人才的抢夺已经到了跨区域、跨国家的范畴。

我们从今年陆陆续续接到很多投资机构、基金、私募形形色色的找到协会，希望从我们的平台上去了解更优秀的大数据项目，企业给予投资等，越来越多的资金关注我们的行业，关注我们的项目。还有一个小事情，我前一阵跟传媒集团的朋

友见面，聊两句突然聊到他要做大数据，传媒行业的数据融合。我去看演出，有一个做演出领域的朋友见面，也是资深的，他提出来要做一个大数据的项目，通过大数据的融合帮助我们的演出观众了解这个了解那个，自动选座等等一些创新。前两年车坏了，去4S修车，很意外被4S店老板请到他的办公室去聊，后来发现他也想做车联网，他也想做大数据。我们给协会的办公室安装了几台德国原装的净化器，因为北京的空气污染很严重，结果厂家的人找到我们说愿意给我们免费安装，前提是能不能帮他们做大数据，基于大数据的客户分析和行为研究。你能看到有大有小，但是大数据很热，我们协会成立了七年，行业筹备、努力了12年，但是我们这两年才感受到开始受宠了，开始有人对我们感兴趣了。

第二个事情，2015年大数据这个行业从另外一个角度看它也很浮躁，现在自媒体发展非常快，微信、微博各种各样的信息，故事满天飞，很多我们的事务所、企业就想了这么多神奇的大数据创造的奇迹，我们中国能不能聊一聊中国的企业创造的奇迹是什么，我们也很想找到这样非常完整的大数据基因的例子，因为这么多的故事，总能找到我们感兴趣的，后来我们找了很多机构，没找到。大家知道我们协会，我们是管理全国三十几个省份的全国数据分析事务所、从业机构，他们每天都在接触企业、案例，都在接触研究，想一想如果我们找不到，恐怕很多人也未必找得到，所以这说明一个什么问题？大数据是很热，但是落地的项目，落地的案例，让我们能感受到中国大数据带来冲击的东西，还不多。



第三个是大数据的会议很多，尤其是这一两年，形形色色，有政府举办的，有企业举办的，有一些地方上的联盟举办的，但是这样的会议很多不是我们真正的分析人员，真正给企业带来价值的人组织的，都是技术流的人组织的，我们在会议上讨论什么是分布式计算，我们的平台、数据化的云也好、存储也好，动不动是几百万、几千万甚至上亿投资的产品，好像大数据是我们用不起的东西，好像大数据是一个需要花很多钱才能构建的东西。其实并不是，只是很多浮躁的东西，未必代表我们真正想看到的本质

说实话数据研究是一个超理性的行业，我们去判断一个企业，去判断一个项目，是按照严谨的逻辑，严谨的决策观点去看的，能不能挣钱，有没有价值。后来我们发现某些大数据技术和大数据产品的企业，他不是按照大数据的思维，是按照更偏向于互联网的思维，这样的企业一方面像打了鸡血一样，拼命往前冲，另一方面我们都能看到他们的团队展示出来的，或者看到他们的焦虑，困兽犹斗那样的一种感觉。所以你们看到其实大数据没那么恐慌，大数据只是做好一件事情，做好一个点，大数据可以给很多人很多企业带来真实的传奇性、有价值的东西。我觉得以我的观点来看，我们不必那样的把大数据

推向浮躁化。

大数据到底是什么？我昨天大会上说过一个观点，业务即数据，数据即业务。今天我们把业务转化为数据，明天我们经营数据就意味着我们经营业务。

所以大数据给我们带来的是真实的企业决策，为企业优化决策、改进决策，不要被大数据过热的东西和过于浮躁的东西所影响。我觉得帮助企业赚钱，使企业平稳发展，这才是大数据给我们带来真实的东西。这里给大家提一些小的意见，建议大家关注企业真实的决策，认真提升自己的数据分析的深度和研究能力，让大数据为企业赚钱，让大数据为企业有效的服务。也给在座参与会议的企业家一点建议，大数据一定要触摸，否则未来你很有可能或者一定会被淘汰，怎么接触它才能最有效？最省钱？最快捷？给大家一个我们在行业中我认为的顺序，第一培养数据分析人才，第二花最少的钱存储你的数据，第三不停的优化分析，让我们企业的的核心研究能力在每一个迭代的过程中不断的提升自己深层次的价值。欢迎大家参与我们的会议，希望大家能通过我们十几场的演讲，给大家带来助力，谢谢大家。FIN

## 【峰会主题演讲精选】大数据人才对零售业未来的影响

编辑 / 中商联数据分析专业委员会市场处 张楠 图 / 崔峻珩

邹东生：非常感谢零售商大会给我这样一个机会，共同聊一聊大数据给我们带来的变化。我今天演讲主题是“大数据人才对零售业未来的影响”。讲三个内容：

- 1、零售业的数据融合
- 2、大数据崛起
- 3、大数据人才对于我们零售行业的价值

应该说现在已经进入了大数据时代，所有的企业必然未来会触碰大数据，无论时间的早晚。那么我们零售行业实际上是最早触碰大数据的，而且也是在所有行业中非常敏感大数据的一个行业，它最主要的原因，是因为零售行业与生俱来具有非常好的大数据基础。我想我们在座很多零售商们，很多年前就已经对自己的数据，企业内部的营运数据，销售数据进行了有效的存储，这些对于我们进入大数据时代以后，都是非常宝贵的财富。

随着零售商行业发展越来越快速，受到了很多电商的冲

击。线下数据已经远远不够为我们做分析，大数据能够帮助我们更好的把线上线下数据加以融合，为零售行业创造越来越多价值。

我们简单聊几个现在零售行业经常使用到的研究方法。第一个是用户画像功能，它是通过对大量的用户数据进行分析，把它进行分类，比如说年龄、性别、文化、收入，还有消费者的喜好。对他进行数据的建模和分析，帮助我们准确的对用户进行定位，进而引导销售。

那么我们也会使用到兴趣图谱，兴趣图谱是把人与人之间共同的兴趣绘画成图谱，分享他们共同的兴趣，找到同类客户相应的核心需求，进而引导我们的零售行业准确地进行营销。

舆情分析，我们可以通过对社交大数据的研究，更好的了解客户对于我们产品各个方面的感受所带来的一些观点、评价、意见，提高客户的购物感受。



动态定价，是通过在线上线下大量的数据，大量的价格数据进行判断现在在国外越来越多零售行业开始使用电子货架标签，这样通过线上线下价格数据的调整，使电子标签的普及，使我们动态定价成为了可能。动态定价可以极大的提升我们的实力。

以上的功能全部是大数据作为支撑，那么现阶段大数据发展到什么样的情况了呢？大数据体量增长变得越来越快，最近两年所产生的数据量已经是人类历史数据的总和，五年以后每天甚至每一个小时产生的数据都可能是之前人类历史数据的总和，大数据已经进入到指数级增长的阶段，数据无所不在。大数据已经成为了美国国家战略，奥巴马说过一句话：大数据是未来的石油。基辛格很早也说过一句话，谁控制了石油，谁控制了世界。可见大数据已经成为了非常重要的战略性能源。

既然大数据很重要，我们很多人也都在接触大数据了，每个人理解都不一样。不同的理解影响企业在大数据时代的布局。有很多人会说，大数据是不是一个概念性的炒作？有没有那么多人说的那么有价值？这是一份报告，里面有一些数据，是麦肯锡在全球研究院做出来的一个研究结果。这是大数据对美国企业的一些影响，里面包括医疗、公共管理、制造业，当然还有很重要是零售业。大数据的应用使美国的企业，零售行业平均利润增长超过了60%，这个数据不是最新的，我相信在美国大数据仍然在蓬勃发展，它给企业带来的影响，会比这个数据还要更多。

大数据为什么会产生？这是因为大数据体量的增长，使原来不可能找到的核心数据成为了可能，使原来不可能分析的决策结果成为可能。所以大数据直接影响到了企业的决策，它可以帮助企业做到更精准的决策，这意味着企业可以承担更低的营运风险，赚更多的钱。所以大数据不需要炒作，因为大数据给企业带来实实在在的价值。

还有一些人会聊到另外一个话题，大数据的核心是不是技术？为什么会有这样一个说法？这几年大家看到大数据越来越热，各式各样的会议，各式各样的活动，很多都是以大数据作为主题的，你会发现一个特点，这类会议的参与者，很多主讲嘉宾都是技术型公司，软件开发公司，聊大数据必然会聊到很多先进的炫目的技术，甚至如果你聊大数据，不聊技术，你就不懂大数据，所以很多人认为大数据是技术。

我们想告诉大家，这个观点是不准确的，或者说是不对的。因为大数据是承接企业最后一公里的服务，它是通过数据的深度分析来完成的，没有分析，只有数据，无法使企业创造价值。而且随着开源结构越来越多，技术构建的费用门槛也会越来越低，甚至将来有些技术会免费。我们这几年也开发了一

套数据化服务平台，虽然我们应用了很多的技术，大量的人力，但是我们会给企业提供公益的，免费的服务。我们想告诉大家，技术没那么重要，什么最重要？深度的分析和对企业未来的预测能力才是不可替代的。所以我想跟大家说，大数据的核心不是技术，大数据的核心是分析能力。

分析人才干嘛，等我有数据我再找人嘛，这个观点并不准确。

有些企业已经开始关注数据存储了，找很多技术公司构建了自己的数据化分析平台，但是这里有一个问题，因为大数据的应用服务于企业的决策，如果你要想更好的做到这个平台的应用，你需要第一件事情先要研究决策方法，研究各种算法，研究各种模型，要去找到数据的核心字段，要去判断数据的深度和广度，要查找互联网数据的来源，只有这样做才有价值。所以要先从分析入手，它可以帮助我们花最少的钱做最多的事情。

第二个，有了数据以后，必须要分析。不分析，没有办法引导决策，所以优秀的分析人才可以帮助企业完成最后一公里的搭建。第三个，优化数据，数据要不断地优化才有价值，模型需要优化，决策方法需要下一个话题聊一聊什么是人才，具有什么样能力的人才是我们所关注的大数据人才？人才，第一个要具备对行业，对业务要熟知，不懂行业，不懂业务，其他具备是没有用的。其次，大数据变革的产生是由于技术的改革，技术的提升，所以懂一定的IT技能，懂一定的技术能力是应该的。再次，大数据不是简简单单的单一学科，它需要你掌握数学、统计、经济学、管理学、决策学等一系列的知识。这些知识可以说数据分析，如果你一定把它作为单独学科，它是一个边缘学科。

这三个能力的交集才形成了真正的数据分析能力。在国外已经把数据分析，人才提到了更高的位置，把它叫做数据科学家。我相信这一类的人才是企业将来必争的人才。

数据化人才将来可能是企业的标配人才，就像企业现在要有会计、财务，数据分析人才将来是企业也必须要具备的人才。而且我们要想跟大家交流一个观点，储备数据人才要早于储备数据和技术。这是因为什么呢？因为大数据人才的应用决定的。首先大数据人才应用有三个方面：数据存储、分析数据、优化数据。有的企业说，我还没有数据，我要先优化，这是需要企业不断地完善自己的研究能力，来提升数据带来的价值。所以我们想告诉大家，数据人才对于未来企业的发展非常重要。FIN

## 关于"CDA注册数据分析师协会"在华运营的调查

文 / 中商联数据分析专业委员会 市场处 编辑 / 李子昂 图 / 崔峻珩

编者按：如果您的认识还停留在“大数据很热很神秘”的阶段，那么您已经OUT了。目前业界对大数据的描述五花八门，但其实大数据应用主要分为两大类：一类是通过大数据来治理国家；另一类就是大数据商业应用，提高商业效益和价值。如今大数据早已走下“神坛”，走入了企业决策实践、走进了国家战略层面。

正是因为大数据行业发展速度之快、需求如此之迫切，才导致了市场上培训机构的日益增多但良莠不齐的现象。作为行业协会，我会有必要对市场不良行为进行管理和约束。为了让广大行业关注者了解真实的数据分析行业、课程情况以及政府对于促进数据分析行业发展的政策举措，特撰写此专题。

广大数据分析行业关注者、爱好者：

近日我会收到会员意见反馈，反映有一家自称在美国成立的“CDA注册数据分析师协会”（以下简称“该机构”）在国内开展业务，并通过北京国富如荷网络科技有限公司（原人大经济论坛，现经管之家）提供“CDA数据分析师”培训及认证。我会对此进行了了解与核实，发现在其运营期间，通过网络及新媒体等方式进行的对外宣传信息中存在着大量的不实言论和夸大宣传，情况说明如下：

一、国内协会在注册成立、运营、监管体系管理很严，

而境外协会不然，特别在一些欧美国家，注册协会只需要提交基本资料及费用即可完成。评价境外协会的实力要通过其在行业内的知名度、影响力，服务体系、会员组成等几方面来进行。我会作为中国数据分析行业唯一的行业主管协会，与亚洲、欧美等国一些知名的从事数据分析和数据管理的行业组织均保持着长期、密切地联系，但从未接触到有关“该机构”的相关资讯。因此，我会在此提示公众，如欲了解境外协会是否真实存在，可以从注册信息、会员组成、行业公信力、国际知名度、学术成果等方面进行核实，而切勿仅凭一个协会的空头



名称来判定他的实力。

二、根据国务院《外国商会管理暂行规定》，境外协会不能在华进行商业性运营。先不论“该机构”是否真实，仅就“该机构”在我国运营所谓的“CDA数据分析师”培训业务并按级别收取培训费、资料费、证书费等各项费用，均与我国法律法规相违背。

三、“该机构”官网上声称注册地址在美国的特拉华州，经我会调查，目前并未核实到相关注册信息，同时“该机构”的全球官方网站是在2014年4月2日才于我国万网注册。由此可见，无论是真实性，还是行业知名度均存在很大的不实

成份。其对外宣称的“海外身份”、“沉淀十年”等言论不由得令人产生质疑。

数据分析的基本从业理念是“客观、真实”，我们希望行业正常健康的发展，我们鼓励正规的培训机构、从业机构更多的进入行业、共同培养大数据人才，但对扰乱行业、虚假操作、欺骗公众、混淆认知的不法机构和行为，我会除进行网络公示以警示公众外，还将收集相关证据、与政府部门配合加强行业监管，维持行业的健康有序发展。FIN

## 协会为CDA数据分析师正名

文 / 中商联数据分析专业委员会 会员处 图 / 崔峻珩

2015年7月中国第三届数据分析峰会成功举办，中国大数据顶尖的企业（百度、阿里、银联智惠）、33个省市的数据分析事务所精英、国内外大数据学者专家齐聚一堂，共同探讨本届峰会的主题——“触摸大数据本质”，3000人的盛会、精彩纷呈的演讲、深入的案例分析使我们久久不能从兴奋回归平静。作为中国大数据行业的主管协会，作为此次峰会的主办方，我们与广大数据分析从业者在共同经历了12年的艰辛历程后，终于看到了大数据研究被社会各界广泛认知和认可，一切的努力变得这么有意义！

在会议的三天中，有个参会者向我询问一件事：是否知道北京有一家论坛，在做CDA的培训，而且在网站上指责协会的CDA人才培养项目是“李鬼”、是假认证。说实话，我实在不知此事，惊讶之余，当时心中颇不以为然，国内培训市场的混乱和无序我是知道的，无非是恶意的宣传和诋毁了。回京后，搜索一下，看到了这个论坛和这篇文章，心中的怒气油然而生，我对恶意的竞争了解而看轻，但对于剽窃中国数据分析行业先行者们十几年的努力、以“创始者”自居，我实在不能释然。既然李鬼已经打上门来，那如果再不为行业正名、不为CDA正名，那就真正愧对行业、愧对那些专业数据分析师们了。

欲正名，先打假！

### 1. “十年磨一剑”之真相

某论坛之宣传：声称自己从05年开始，沉淀了10年的数据分析研究，是数据分析人才的领军机构，率先推出数据分析

培训体系。

假在何处：

该论坛的实体运营公司是06年成立的，先不论是否成立就开始数据分析培训，06年至今也不足10年。

论坛虽有大量的发帖，但如果查询发帖历史，会发现2010年前几乎没有任何关于数据分析的帖子，那么10年的分析沉淀，600万学习资料及培训体系由何得来？

真相是什么：

中国数据分析行业以2003年为我国的大数据元年，这一年的9月，原信息产业部（现工信部）公布了数据分析人才培养计划，中国开始了第一批数据分析人才的培养工作，到2005年有了第一家项目数据分析师事务所，2008年全国行业协会成立（中国商业联合会数据分析专业委员会，以下简称“协会”），目前全国已有33个省市成立了项目数据分析师事务所，协会拥有几万名数据分析人才，500多名执业数据分析师，随着“大数据时代”来临，行业影响力越来越大。这才是真正的十年磨一剑！大量的历史资料，请大家自己验证。

协会的数据分析人才培养逐年加深，数据分析培训体系已完成了5版的全面改进，2015年4月，协会公布了最新版（第5版）的课程体系。培训体系的完善是在大量数据分析专业人才和专家共同付出无数努力后成就的，不是吹吹牛就可以称为体系的！

### 2. 主办机构及认证之真相

某论坛之宣传：某论坛CDA数据分析研究院发表严正声

明，提出自己是正版，别人为山寨！指出证书无用，技能更强。

假在何处：

主办机构何在？某论坛是主办机构吗？论坛本就是虚拟网站，不具备法人资质，无法作为主办机构。那所谓的CDA数据分析研究院也并未公开其营业执照，其真实度与合法性大大存疑。

仔细看这个网站，发现对方还有一个CDA数据分析协会的证书，真让人费解，如果有这么正规的协会，为什么遮遮掩掩，只说CDA数据分析研究院，行业协会的认证不是更有说服力吗？

如果证书无用，为什么又起个“数据分析师”的名字？干脆称：数据分析培训，不好吗？如果证书真的没用，何必又要通过一个无法确认真实性的海外协会来颁发证书？

真相是什么：

所谓的CDA注册数据分析师协会，号称是在美国注册的协会，我们先不谈协会声名真实与否，即使是真实知名的境外协会，在中国范围内直接经营也是不合法的，可参阅国务院颁发的《外国商会管理暂行规定》；另外，美国CDA注册数据分析师协会有会员吗？全美国又有多少CDA证书持有者呢？事实上这家协会的官方网站是2014年4月刚刚注册的，在海外毫无任何知名度可言，只是作为他们虚假宣传的掩护。

中国商业联合会数据分析专业委员会是接受国资委监管、民政部正式批复的，于2008年4月成立的全国性行业协会，行使行业管理和服务的权利，是中国数据分析行业的全国性协会，欢迎大家官网上通过批复文件及营业执照了解真伪。

协会的CDA项目的证书是由工信部教育与考试中心及主管协会分别颁发，是中国数据分析行业及社会广泛认可的证书，为学员今后进入行业，取得从业地位奠定了保障。

### 3. 学习是一种气质之真相

某论坛之宣传：该论坛CDA认证培训，体现出学习是一种气质，是解决实际问题、掌握真本事的培训，是让人心舒适的培训。

假在何处：

说法没错，学习是一种气质，无论什么样的培训认证，都要以学到真本领为真正目标，可这样一个大肆虚假宣传，建立“空壳”海外协会，为学员颁发无公证力的证书，又如何为学员提供真正的培训保障，让学员学到真本事？宣传中号称的数据分析培训十年，又真正培养了多少分析师在企业中从业呢，可以露露脸吗？

不说面授学习，远程学习的费用居然每人从4400 - 13000

元不等。既然说证书没用，为什么还需要学员额外花1000或1500元一次的考试费去取得一张无知名度的证书？这只是通过收取高额远程教学学费和额外考试费，从学员身上多次牟利罢了。

真相是什么：

协会与工信部教育与考试中心共同主持的CDA（数据分析师）人才培养项目，主旨是：为中国培养基础数据分析专业人才，满足“大数据时代”企业对数据分析人才的需求。

数据分析人才培养项目已开展12年之久，全国2万名数据分析人才分布全国各地，学生对课程的满意度平均在90%以上，由于行业协会管理着全国的从业数据分析事务所，用他们遇到的真实难题与案例反哺和滋养课程，并广泛对接企业大数据分析的需求，使课程的实战性很强，学以致用是行业培养人才的基础理念。

协会的CDA数据分析师培训，也是采取远程教育的方式，聘请行业内执业分析师作为师资精英授课，课程生动而实用，培训及考核全部费用仅每人1300元。在课程结束通过考试后，工信部教育与考试中心会为学员颁发数据分析师证书，协会也会将学员信息存入数据分析人才库并颁发数据分析师人才信息证明，为学员从业铺平道路。

为促进行业发展，让越来越多的人掌握数据分析技能，协会在全国各地定期举办丰富的公益沙龙和俱乐部活动，频次高、案例讨论多，并聘请行业内著名数据分析师作为讲师，免费邀请全国学员参与。此外，协会以向学员免费派发《中国数据分析》会员特刊、跟踪课程更新、以及年检制度和针对性培训等方式方法为学员提供终身后续培训服务。

作为数据分析行业主管机构，我们支持整个行业的发展，支持正规培训机构开展数据分析培训服务，为学员提供更多的服务选择。但不良培训机构开始混淆市场，妄图用劣币驱逐良币；又借着不知名的海外协会身份在数据分析培训市场上大肆虚假宣传，欺骗学员，从学员身上利用分级制度诱导学员缴纳多次学费参与多次学习，并通过收取考试、材料费等行为反复牟利。

如果此时我们作为行业监管机构再默不做声，并冷眼旁观这样的行为，让更多学员上当受骗，那岂不是天大的失职？所以必须打假，为真正的CDA培训服务正名！也希望广大关注“大数据”、关注“数据分析”的学员们，可以理性看待培训、理性看待行业，不要被虚假的包装和精密的话术所迷惑，最后做出理性的决断。FIN



## 经济日报：想考CDA数据分析师，需擦亮眼睛

文 / 中商联数据分析专业委员会 会员处 编辑 / 李子昂 图 / 崔峻珩

编者按：7月30日，经济日报在其微博、微信公众号上发表文章《10张最值得考的经管类证书》，经我会监测，其中含有不实信息。随后，经济日报对此进行了核查，并对此文予以删除。



出于对大数据发展的关注，经济日报对我会副秘书长蒋文伟先生进行了采访，全文如下：

随着大数据的发展，数据分析行业与相关工作岗位开始逐渐受到各大企业重视。而在此行业发展人才紧缺的当口下，数据分析行业培训认证工作中也出现了鱼龙混杂的情况，一些不正规的培训机构开始运用新媒体进行虚假宣传，对学员形成不良影响。为了更加了解行业状况，辨析真伪，我们联系到了我国数据分析行业唯一的主管行业组织——中国商业联合会数据分析专业委员会（以下简称：“协会”）蒋文伟副秘书长，对行业人才培养的情况有了进一步了解。

蒋秘书长指出，现在中国正处于一个大数据兴起和高速发展的时代，在整个行业市场还不成熟及有待规范的情况下，一些培训机构看到数据分析培训市场的发展空间，为了短期牟利，采取不正当竞争及夸大宣传等行为，误导想要学习数据分析知识的学员。

劣质的课程和师资、昂贵的学费，不仅无法使学员学到真正的数据分析知识和技能，而且也使社会各界对大数据人才的价值产生了误解，这些都不利于行业的快速发展。由此，协会已针对此种情况研究对策，对明确扰乱市场环境、欺骗学员的数据分析培训机构进行网络公示，向社会警示风险。蒋秘书长说：数据分析的基本从业理念即是：客观、真实，我们鼓励正规的培训机构更多的进入行业，共同培养大数据人才，但虚

假的证书、虚假的宣传，毁坏的是整个行业的口碑。

据悉，数据分析作为一个独立行业在中国开始发展，是从2003年CPDA项目数据分析师培养计划开始的，至今已发展了12年之久。2013年“大数据”概念的引爆，使得政府、企业和个人更为广泛地给予了这个行业以关注，也正是由此，各行各业对懂“数据”、精“分析”的专业人才产生了迫切需求。因为只有培养和储备数据分析人才，才能通过数据人才对数据的分析引

导决策，帮助企业完成最后的搭建。因此一个对行业熟知、懂IT技能、掌握数学、统计、管理学等一系列知识的人才成为企业必须要具备的标配人才。可以说，数据分析人才的培养是推动整个数据分析行业发展的重中之重，也是大数据时代企业盈利的根源所在。

同时，蒋秘书长也表明，当前行业中基础数据分析人才仍存在很大缺口，以后随着数据分析行业的进一步发展，将会有越来越多的企业对基础数据分析岗位产生需求。因此，本着促进行业发展，为行业输送合格的数据分析基础人才的目标，协会与工信部教育与考试中心联合推出基础型数据分析人才培养计划——CDA数据分析师。

笔者提醒广大数据分析爱好者，在培训认证市场较混乱的今天，请用理性思维去看待行业发展，同时提高辨识能力，切莫被虚假或者夸张宣传而误导。在选择数据行业培训机构时，应当证实一下培训机构是否有运营资质、证书是否由国家或行业正规机构颁发及认可。

备注：中国商业联合会数据分析专业委员会成立于2008年4月，是经国务院国有资产监督管理委员会审核同意、中华人民共和国民政部正式批准和登记的中国数据分析行业全国性协会组织。协会自成立以来，一直在项目引进、知识普及、行业监督、数据分析人才培养及专业事务所建立方面发挥着积极地促进作用。FIN



## 大数据时代，技术和分析哪个更重要？

文 / 中商联数据分析专业委员会 会员处 编辑 / 苏然 图 / 崔峻珩

资深数据分析师，戴文波特在《哈佛商业评论》上的撰文《数据分析师的崛起》中提到，大数据时代的到来意味着处理庞大的数据将会在每个人的工作中，占有越来越大的比重。因此，对经理人和员工来说，数据分析和数据认知能力将变得无比重要。

此外，《埃维诺调查》中的一次报告结果显示，超过百分之六十的管理者认为他们的员工需要提高分析能力去将数据转化

为洞察力和商业价值。许多行业的高管都已注意到了数据分析的重要性，并认为数据分析能力及数据分析人才是企业发展的必需品。而在真正通过数据为企业做出贡献这个问题上，分析，很多时候比数据本身更重要。

数据是一种资源，并且是一种可以重复使用和共享的资源。但是数据如果不能被以正确的方式使用，就只会成为一种负载。数据本身并不产生价值，只有通过合理分析与使用，才能产生和体现它的价值，而不是因为采用了顶尖的技术和算法才体现了它的价值。

所以，如何分析和利用数据对业务产生帮助才是关键。分析的过程就显得尤其重要，只有用好分析能力，才能通过数据来解决真实问题。在这当中，技术仅仅是数据分析中的一部分，只是工具与手段。而企业只有将分析方法嵌入到由分析师参与的商业决策中，才能真正体现出数据的价值。而在具体通过数据分析体现数据价值的过程中，模型算法的设立与预期目标正是由人的思维理念和分析能力决定的。

无论是SAS, R, Matlab, Excel, Hadoop还是SPSS, 都只是重要的数据工具，如果只掌握工具的用法，对于分析师在企业中解决实际问题没有任何帮助。因为工具与技术只是解决问题的手段，而真正在企业项目实施中起到决定性作用的，是分析能力。假如一名数据分析师，只重视技术手段，而缺乏分析能力的话，就只能成为工具的使用者，而不能成为数据的掌控者，更不能成为解决问题的分析者。

随着大数据时代的发展，数据基础设备的进一步完善，



越来越多的数据得以被发掘、利用，而随着技术的不断革新，数据分析工作者便可以更快更准确的得到结果。但在企业中遇到的真实问题与案例是复杂的，不同岗位的员工与客户从数据中解读到的信息都是不同的，因此，最后得出的结果也会不一样。

当分析师们为一个商业问题提供数据分析解决方案时，需要将客户的需求与情况作为分析的基础并随时进行调整。只有这

样，才能真正取得详细的，准确的报告与方案。如果只会用技术，而在具体分析能力上有欠缺的话，就只能按照流程，通过模式化的数据清洗与建模来得到结果。这就会缺少具体情况的融入与分析思想的引导，只能是依葫芦画瓢，和实际情况出入巨大。由此可以看出，在数据分析中，技术只是一个环节，而分析才是一切的核心。

### 总结

在大数据时代的浪潮中，数据存储量将会数以亿计的发展，膨胀。随着大数据分析智能平台的诞生与应用，分析工具与技术将进一步简化，这意味着大数据分析的技术门槛将进一步降低，而真正体现结果差距的，正是分析能力与方法。通过合理的分析，企业可以从这如此浩瀚的数据中找到自己的决策方向，找到自己的发展远景。

每位企业员工也可以通过提升个人的数据分析能力去驾驭、管理数据并借此提升自己的水平。而在学习数据分析的过程中，要分清主次，选择案例分析出色，更重视培养自身分析能力的课程。也一定要明白，一切数据分析的根基，根虽然在数据，但是一切的基础却是分析。FIN





## 为什么CPDA项目数据分析师不分级？

文 / 中商联数据分析专业委员会 会员处 编辑 / 李子昂 图 / 崔峻珩

CPDA项目数据分析师至今已开展12年之久，培养了几万名分析师，成立了逾百家项目数据分析师事务所，遍布在全国各地，渗透于各行各业。这些成就既是行业内数据分析专家的辛勤努力滋养CPDA课程造就的，也是这个大数据时代的趋势促成的。有学员一直在咨询协会这样一个问题：为什么CPDA项目数据分析师不去区分等级呢？

仔细想想，这是个好问题。那么究竟为什么CPDA项目数据分析师认证不去分什么初级、中级、高级，亦或一级、二级、三级呢？

国际惯例，建行业标准

事实上，国际上知名的、专业的从业认证体系都不分级别，例如：注册会计师（CPA）、特许公认会计师（ACCA）和注册税务师（CTA）等等。不仅如此，而且此类证书，不是终身有效的，而是持有者需通过定期年检以保持其资质的有效。注册税务师是三年一次年检外，注册会计师和特许公认会计师均为一年一检。

CPDA项目数据分析师依照国际从业认证的规范，从2003年设立开始，即不设立任何级别，每三年年检一次。学员只有

通过严格考核并合格后，才能拿到由行业协会颁发的项目数据分析师证书，取得从业资质。坚持CPDA项目数据分析师不区分等级，正是协会严格尊重国际惯例、遵守从业规范的结果。

从业与职业

您也许会想到，虽然国际上顶尖的从业认证都不分级别，但同样国际知名的特许金融分析师（CFA）可是分了三个等级的啊。要想明白这个问题，就得知道从业与职业的区别。特许金融分析师是一个职业能力证明，是对人员技术和知识的证明，并不涉及一个行业及第三方服务的从业能力。而拥有从业认证的人才可以通过成立中介或事务所来提供服务。如注册会计师（会计师事务所），执业律师（律师事务所）和项目数据分析师（项目数据分析师事务所）。

CPDA项目数据分析师是具有从业能力的数据分析人才，中国数据分析行业的执业人才，具有从业报告的签章权，可成立项目数据分析师事务所，通过提供专业的第三方中介服务为各个行业提升业务经营水平。

执业证书，不是上岗证

中国的职业资格体系，是国家劳动人事部职业资格鉴定

指导中心开设的职业资格认证，俗称的“上岗证”，普遍分为员、助理师、师、高级师等级别。虽然证书被冠以“职业资格”，但普遍存在课程水平不高、专业性低、社会认可低、培训市场混乱等问题，分级方式使学员重复学习并考级，与行业的从业模式相悖。迄今为止，国务院已经取消了200多项职业资格，并会进一步治理整顿。

CPDA项目数据分析师是中国数据分析的执业人才的组成部分，非职业认证体系，自然也不必分级，专业协会颁发的专业证书，未来含金量将进一步提升。

只为学本事，不为评职称

职称作为中国现状下主要职位和社会地位的代表，较高职称的人享有较高薪酬与福利待遇，但与岗位实际技能并未必有直接的关联。但随着职业资格的逐渐取消，可以预见到的是职称体系也即将在中国社会的进一步发展过程中逐渐过渡。而CPDA项目数据分析师之所以不区分等级的原因，一是协会并不愿意为数据分析这样一个重视新技术与新思想的行业建立较为僵化的职称标准；二是协会本着学以致用、学以致用的行业育人理念，期望每位数据分析从业人才将更多的精力放在技术结合实践上，而不是一直疲于应付各类资格考试。此外，CPDA项目数据分析师作为数据分析行业的从业认证，协会更注重培养从业人员的能力与技术，更加真实的推动行业发展，而不是以职称

论英雄。

信守行业准则，坚持从业理念

现在数据分析市场上有培训产品打着“海外”旗号大肆虚假宣传，推出培训分级证书，更多是为了诱导学员重复消费和学习而采取的牟利手段，通过将课程进行无意义的划分等级诱使学生一级一级的去缴费学习、付费考试。

CPDA项目数据分析师已发展了12年，一直坚守行业准则，从未考虑将证书分级，为学员和分析师负责也就是对行业发展负责！

促行业发展，育基础人才（CPDA与CDA的区别和联系）

随着大数据时代的兴起与发展，基础数据分析人才岗位产生大量需求。为了向企业输送基础数据分析人才，填补企业基础数据分析人才漏洞，解决学生工作经验缺乏及用人单位的困境，协会推出了面向基础分析人才的培训项目——CDA数据分析师。CDA数据分析师是对基础数据分析人才的培养计划，两者之间并不是分级或层级关系。

CPDA项目数据分析师是中国数据分析的从业人才，可以组建项目数据分析师事务所并成为协会的执业会员。两者服务人群与目标皆不同，无任何等级关系。FIN

## 国务院印发《促进大数据发展行动纲要》

文 / 中商联数据分析专业委员会 会员处 编辑 / 张楠 图 / 崔峻珩

新华社北京9月5日电 经李克强总理签批，国务院日前印发《促进大数据发展行动纲要》（以下简称《纲要》），系统部署大数据发展工作。

信息技术与经济社会的交汇融合引发了数据迅猛增长，数据已成为国家基础性战略资源。坚持创新驱动发展，加快大数据部署，深化大数据应用，已成为稳增长、促改革、调结构、惠民生和推动政府治理能力现代化的内在需要和必然选择。

《纲要》提出，要加强顶层设计和统筹协调，大力推动政府信息系统和公共数据互联开放共享，加快政府信息平台整合，消除信息孤岛，推进数据资源向社会开放，增强政府公信力，引导社会发展，服务公众企业；以企业为主体，营造宽松公平环境，加大大数据关键技术研发、产业发展和人才培养力



度，着力推进数据汇集和发掘，深化大数据在各行业创新应用，促进大数据产业健康发展；完善法规制度和标准体系，科学规范利用大数据，切实保障数据安全。《纲要》明确，推动大数据发展和应用，在未来5至10年打造精准治理、多方协作



的社会治理新模式，建立运行平稳、安全高效的经济运行新机制，构建以人为本、惠及全民的民生服务新体系，开启大众创业、万众创新创新驱动新格局，培育高端智能、新兴繁荣的产业发展新生态。

《纲要》部署三方面主要任务。一要加快政府数据开放共享，推动资源整合，提升治理能力。大力推动政府部门数据共享，稳步推动公共数据资源开放，统筹规划大数据基础设施建设，支持宏观调控科学化，推动政府治理精准化，推进商事服务便捷化，促进安全保障高效化，加快民生服务普惠化。二要推动产业创新发展，培育新业态，助力经济转型。发展大数据在工业、新兴产业、农业农村等行业领域应用，推动大数据发展与科技创新有机结合，推进基础研究和核心技术攻关，形成大数据产品体系，完善大数据产业链。三要强化安全保障，提高管理水平，促进健康发展。健全大数据安全保障体

系，强化安全支撑。《纲要》还明确七方面政策机制。一是建立国家大数据发展和应用统筹协调机制。二是加快法规制度建设，积极研究数据开放、保护等方面制度。三是健全市场发展机制，鼓励政府与企业、社会机构开展合作。四是建立标准规范体系，积极参与相关国际标准制定工作。五是加大财政金融支持，推动建设一批国际领先的重大示范工程。六是加强专业人才培养，建立健全多层次、多类型的大数据人才培养体系。七是促进国际交流合作，建立完善国际合作机制。

《纲要》要求，各有关部门要进一步统一思想，认真落实各项任务，共同推动形成公共信息资源共享共用和大数据产业健康发展的良好格局。

FIN

## 工信部正制定《大数据产业“十三五”发展规划》

文 / 新华社记者 赵晓辉 编辑 / 李子昂 图 / 崔峻珩

新华社北京9月10日电（记者 赵晓辉）工信部信息化和软件服务业司司长陈伟10日在接受记者采访时说，工信部正在制定《大数据产业“十三五”发展规划》，支持大数据技术和产业创新发展，提升大产业支撑能力，培育新业态新模式。

近日，国务院印发《促进大数据发展行动纲要》，这为我国大数据发展进行了顶层设计和统筹部署，工信部主要负责大数据产业发展以及应用示范相关工作。陈伟说，除制定《大数据产业“十三五”发展规划》外，工信部还将出台促进大数据产业发展的推进计划，促进规划、标准、技术、产业、安全、应用的协同发展。

据介绍，工信部将组织实施“大数据关键技术及产品研发与产业化工程”，通过相关项目和资金引导和支持关键技术研发及产业化，同时开发面向工业、电信、金融、交通、医疗等数据密集型行业的大数据应用解决方案。

“要促进大数据与其他产业的融合发展，着力发展工业大数据，加强产业生态体系建设。”陈伟说。

他说，围绕落实中国制造2025，支持开发工业大数据解决方案，利用大数据培育发展制造业新业态，开展工业大数据创新应用试点。同时，促进大数据、云计算、工业互联网、3D打印、个性化定制等的融合集成，推动制造模式变革和工

业转型升级。

据介绍，目前，工信部已经指导全国信息技术标准化委员会组建由130余家单位构成的大数

据标准工作组，组织起草了《大数据标准化白皮书》，制定大数据标准体系，已经开展数据质量、数据安全、数据开放共享和交易等方面的多项国家标准的立项和研制工作。

在安全保障方面，陈伟说，未来要研究制定网络数据采集、传输、存储、使用管理的标准规范，加大对隐私信息保护、网络安全保障、跨境数据流动的管理，组织开展相关的专项检查和治理。

他表示，要推动和配合相关部门组织开展数据共享、开放、交易、安全等方面的立法研究工作，解决制约大数据产业发展体制机制因素和不确定性的市场因素，为产业和应用发展营造良好法规和市场环境。FIN





## 新兴项目数据分析师事务所如何走好创业之路

文 / 《中国数据分析》 记者 张明珠 图 / 崔峻珩

睿智愈发成为当前时代的代名词，而睿智的决策往往离不开精密的数据分析。企业的投资决策也愈加与数据分析紧密相连，“拍脑门”的决策方式已经是过去式。随着数据分析的需求越来越迫切，在企业与专业的数据分析之间，催生出了这样一个机构：项目数据分析师事务所（以下简称事务所）。

随着市场需求越来越大，全国各地的事务所数量也随之增加。为此，2015年8月，中国商业联合会数据分析专业委员会会长邹东生先生对一些地方事务所进行了实地考察。考察的对象主要是一些刚成立的新兴事务所。

本期会客厅，我们就将针对邹会长的考察情况，对他进行一次深入的访谈。旨在通过本次采访，了解新兴的事务所该如何走好创业之路。

本刊记者：邹会长，很高兴您能够抽时间接受我们的专访，首先我想了解下，您认为新兴事务所进入的难点有哪些？

邹东生：第一是场地：实际上，中介服务机构对场地的要求较高，是对外服务的形象。办公场地的规模不需要很大，

但是一定要保障办公环境的品质，否则会影响客户的印象，影响落单；第二是人员：最难的不在于业务开展，而是在于专业能力的提升。随着大数据在中国不断升温，基于大数据的业务需求在迅速增加，但大数据对于团队专业能力的要求是较高的，只有这支团队给出的解决方案真正能打动客户，需求才能转换成真正的收益。有的事务所对于专业的理解不够，认为如果业务来了再临时抓人或攒个方案就能抓住客户，这种投机心理只能使事务所的腾飞成为“镜中月水中花”；第三是团队：Team Leader 必须了解团队每个成员，了解他们每个人对数据分析的职业规划，帮助每个成员了解大数据研究的真正核心，认清从业方向。

本刊记者：您在与这些新兴事务所团队的沟通过程中发现了哪些问题？

邹东生：在这次交流过程中，我也跟这些地方新兴事务所的团队聊了很多关于从业思路和方向的问题，对他们从前的认知是一个颠覆。比如，他们的认知存在几个普遍的误区：第一，他们认为数据分析工作很像管理咨询；第二，数据分析的



研究深度远远比不过管理咨询；第三，数据分析是一个更技术化的工作，这些都在数据分析的认知方面有较大的偏差。更多了解大数据发展的背景、更多阅读数据分析的专业案例或研究方案，会使分析师团队树立正确的从业方向和理念，试想一下，如果分析师对自己的职业都没有深入理解和认识，无法真正了解大数据的魅力，那么又如何能让客户认可我们的服务呢？所以，协会在近期会对刚刚成立的事务所给予更多的行业引导方面的帮助和支持，让新兴事务所团队少走弯路。

本刊记者：邹会长，协会对于新事务所入会的审核标准是什么呢？

邹东生：一个新兴事务所不管基础怎样，只要抱着对数据分析行高度的热情和全力以赴的投入，它就有加入行业的机会，因为它本身不属于一个资本型产业，而是一个智力型产业。我们为什么要制定这样一个入会评审体系，目的第一就是要杜绝事务所在成立的时候抱着一种投机心态。像一些早期的事务所，成立不是为了在数据分析行业努力的打拼，而是为了拥有资质后，坐等业务上门，只是想在新兴市场上占有一个位子。对于这种投机心理偏重的事务所，协会是不会接纳的。第二就是准备过度不充分的，对于市场的困难、挑战几乎没有准备，完全带着一腔热情，对于这样的事务所，我们建议它暂缓入会。毕竟协会希望培养更多专业型的事务所，并不希望我们的事务所过多和过度分散，宁缺毋滥。

本刊记者：您给这些创业事务所的建议有哪些？

邹东生：通过以上也说明了一个问题，不管大数据也好，互联网+也好，任何一种先进的技术都要从一板一眼的品牌建立、品牌推广开始。资本是用来助推业务的，但不能被资本所绑架。新兴事务所团队需要的是一个实实在在做项目的心态。从协会方面来讲，作为主管协会我们也会给新兴事务所提一些建议，让其少走弯路。数据分析是处在风口的行业、最有价值的行业。但是这个行业的门槛是非常高的，它要求有比较强的专业素养和能力，不意味着是CPDA就一定能创业成功，所以我给我们的分析师一个建议：成为项目数据分析师说明你已经正式进入数据分析行业，具备了基本的数据分析能力，但作为一名分析师要有终身学习的准备，不断的深造才是成功的必要条件。另外，作为一个咨询性质的行业，跟客户进行洽谈的时候，需要做一些准备，比如，事务所的介绍、开展的业务范围、成功案例展示等；在推广模式上，网站上的工作也要做足，互联网处在一个高速发展的阶段，推广过程中没有网站，这就意味着宣传过程中一个重要窗口的缺失，对于一个新成立的事务所来讲，网站建设和优化十分重要。

本刊记者：对于新兴事务所的业务方向您有什么建议？

邹东生：新兴的事务所在刚进入的时候会有许多问号和困难，最初是对这个行业的憧憬，但是越往前走他们会面临无数个“十字路口”。首先，业务方向还是要偏向量化经营，大数据毕竟正处于一个初级阶段，这个阶段的业务特征是技术咨询模块在发挥作用，再具体说，最大的一个业务亮点就是企业的数据化。在企业将来的业务层面上，不管是做营销还是别的，任何一个企业都需要对自己的数据进行存储、都需要专业的数据分析知识，这就意味着巨大的企业需求。现阶段在战略性模块下，企业的数据化是企业未来竞争的必经之路。在这个过程中，对于一个新兴事务所来讲，做数据化和平台搭建是一个非常重要的盈利模式，也是现阶段最容易的赚钱机会。

本刊记者：那么业务收费从何而来呢？

邹东生：第一，咨询服务费用。因为事务所跟客户谈业务时，不是上来就介绍产品，要摆脱技术模块的束缚。事务所不是一个软件公司，而是一个咨询公司，所以一定要收咨询服务费用；第二，要收平台的构建费用，根据模块收费；第三，算法的优化及长期服务费用，这也是体现事务所长期服务价值之所在。构建这样一个数据化平台，不仅是帮助事务所赚到今天数据化服务的钱，而且是构建了事务所和企业之间长期稳定合作的构架关系，真正为企业做好长期服务。

总结：此次访谈，根据邹东生会长到一些新兴事务所实地考察的结果，针对新兴事务所的创业和业务起步阶段应注意的问题给出了建议。总结为以下几点：

- 1、新兴事务所要保障办公场地的品质，这关乎到客户对事务所的第一印象，也影响着落单率。
- 2、团队的领导人要了解团队中每个成员的从业方向和技能掌握情况，善于用人，发挥个人所长，凝结团队成员力量，提升业务水平。
- 3、深入理解行业特质，避免行业认知误区。
- 4、从协会角度来讲，我会会员处将尽量给予事务所支持与帮助，给出合理建议，避免新兴事务所的创业之路受阻。
- 5、协会有严格的入会审核制度，对于抱有投机心态、准备过度不充分、开展不良业务的新兴事务所申请入会的，协会建议暂缓入会或拒绝接纳。FIN

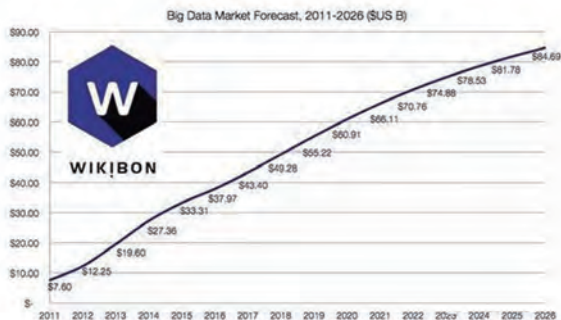
# 2015年大数据展望和市场预测

文 / 大数据文摘 编译 / 丁雪 校对 / 姚佳灵 席雄芬 图 / 崔峻珩

本文中提到了在研究中使用的不同方法，许多方法得出了相同的结论。更为深入地洞察顾客及其需求是优先考虑的事，在如何优化销售周期及精简客户服务上获得更精确的信息也是经常要考虑的。最成功的大数据应用案例为我们展示了企业是如何突破限制变得更加关注和响应顾客的需求。

以下给出的是最近的展望与预测综述：

wikibon预计大数据市场从2011年到2026年将获得17%年复合增长率，将在2026年达到840亿美元的高峰。大数据市场从2013年的196亿美元增至2014年的273.6亿美元。这些和其它分析结果来自wikibon关于大数据市场普及和增长非常棒的研究。下图展示了他们的大数据市场预测概况。



根据Forrester Wave最新的报告《Big Data Predictive Analytics Solutions, Q22015》，IBM和SAS是大数据预测分析市场的领军企业。最新的Forrester Wave报告是基于13家不同的大数据预测分析提供商所提供数据的一个分析，这些

公司分别是AlpineDataLabs, Alteryx, AngossSoftware, Dell, FICO, IBM, KNIME.com, Microsoft, Oracle, Predixion Software, RapidMiner, SAP, andSAS。Forrester特别指出Microsoft Azure Learning是大数据分析市场令人印象深刻的新的新贵，它显示了微软公司作为突出的参与者在大数据分析市场巨大的潜力。

Figure 3 The Forrester Wave™: Big Data Predictive Analytics Solutions, Q2'15



根据高德纳咨询公司最新的魔力象限（Magic Quadrant）分析，IBM、KNIME、RapidMiner和SAS在高端大数据分析平台市场遥遥领先。

在其最新的报告中，评估了16家领先的大数据分析平台提供商从草案到构建解决方案的分析技术。报告中包括了以下的供应商：AlpineDataLabs, Alteryx, Angoss, Dell, FICO, IBM, KNIME, Microsoft, Predixion, Prognoz, RapidMiner, RevolutionA analytics, SalfordSystems, SAP, SAS和Tibco Software。Gregory



Piatetsky将魔力象限 ( Magic Quadrant ) 的关于大数据高端分析平台的调查结果转化为排名的形式，并提出自己的见解。



Salesforce公司预测在2014财年通过增加数据分析和商业智能 ( BI ) 应用将会增加该公司的潜在市场范围到130亿美元。该公司计划在2018年前通过增加新的分析应用，将潜在市场范围扩大至820亿美元。从2013年到2018年期间，力争在潜在市场范围获得11%的复合增长率。



正如互联网对商业运营带来的深刻影响，89%的商业领袖相信大数据也将带来革命。

85%的商业领袖坚信大数据会给商业运营带来翻天覆地的变化。79%的商业领袖同意这么一种说法：“那些不接受

大数据的企业将失去它们的竞争地位，甚至面临被淘汰的可能。” 83%的商业领袖公司已经开始追求大数据项目以夺取新的竞争优势。企业中受到大数据影响的前三大领域是：影响客户关系 ( 37% ) ；重新定义产品开发 ( 26% ) ；改变企业运营方式 ( 15% ) 。下图对比了在未来五年中在企业里受到大数据影响最大的六个业务领域。



全球大数据市场预计在2025年将达到1220亿美元的收益。同时，弗若斯特沙利文咨询公司 ( Frost&Sullivan ) 预测到2025年全球数据流量每年将会达到100泽字节。

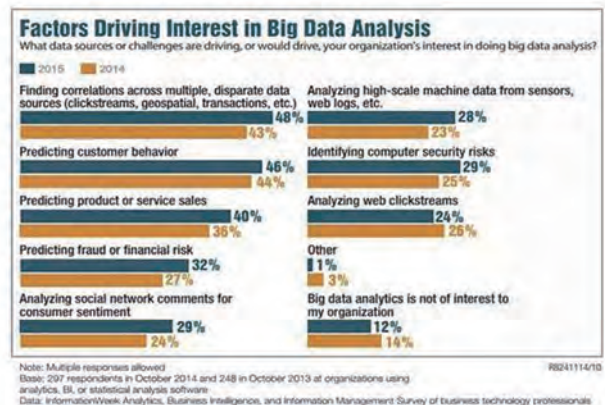
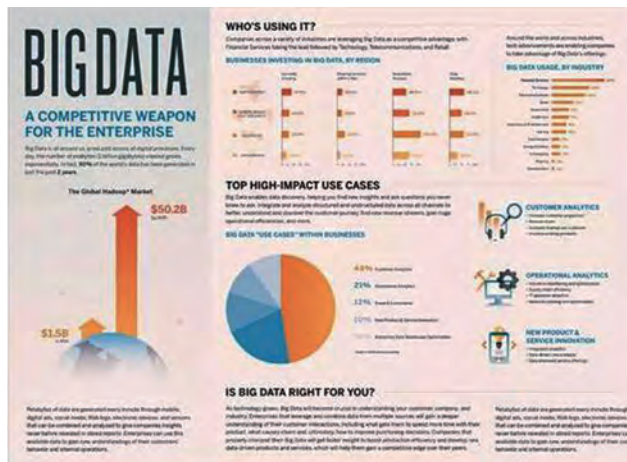


从2014年到2020年，全球文本分析市场的复合年增长率 ( CAGR ) 达到25.2%，市场价值有望于2020年达到65亿美元。客户关系管理 ( CRM ) 、预测分析和品牌信誉营销是排名前三的项目应用。下图展示了报告中关键研究结果的概述。



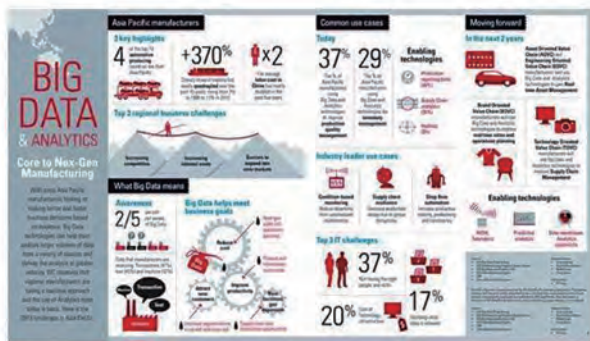
客户分析（48%）、业务分析（21%）和商业欺诈与商业尽职调查（21%）是大数据应用最为广泛的三个案例。Datameer对市场分析后发现，全球Hadoop市场有望由2012年的15亿美元增至2020年的502亿美元。金融服务、技术和通讯是如今运用大数据解决方案的三个领先行业。

发现不同数据源中的相关性（48%）、预测客户行为（46%）和预测产品与服务销售(40%)这三个主要因素推动了大数据分析兴趣的增加。这些以及其它来自2015年《信息周刊》和《商业智能调查报告》中有趣的发现，让我们窥见了企业是如何选择分析应用和平台。



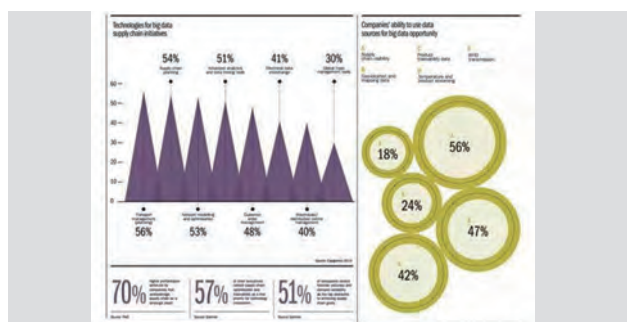
亚太地区37%的制造商正在借助大数据和分析技术提高生产质量管理。国际数据公司(IDC)发现这一地区的制造商更依赖于上述技术来降低成本、提高生产力和吸引新客户。

在过去的一年中，银行、通讯业、媒体、公共事业和批发贸易都大大加强了大数据分析的应用。高德纳咨询公司针对2014年第四季度大数据应用作了最新调查，调查发现基于柱状/内存数据库技术、大容量数据仓库技术，基于索引的检索技术和日志数据分析技术在相关文献报告中的引用和使用有超过5%的增长。



国际数据公司(IDC)预测今年全球大数据与分析技术市场在硬件、软件和服务上达到1250亿美元的总收益。由于日益复杂的系统配置和企业级的全面解决方案，专业的服务/技术的比率今年也会增长25%。国际数据公司(IDC)也预测物联网（IoT）市场从2014年到2019年复合年增长率将达到30%。FIN

在供应链管理中，供应链可视化（56%）、地理位置和地图数据（47%）以及产品可追溯数据（42%）是大数据获得机会的三大潜在领域。运输管理、供应链规划、网络的建模和优化是供应链方案中最热门的大数据应用。





BIG DATA CYCLE  
 BILLION  
 PETABYTES  
 PROCESS  
 USING TIME  
 TARGET  
 HUNDREDS  
 DISK  
 BUSINESS  
 NETWORK  
 SYSTEM  
 SOFTWARE  
 VOLUME  
 ORGANIZE  
 RESEARCH  
 MASSIVELY  
 GOVERNANCE  
 ANALYSIS  
 CLOUD BASED  
 STORAGE  
 CAPTURE  
 OPTIMIZE  
 SUPPORT  
 MPP  
 OPTIMIZE  
 MANEAGEMENT  
 TECHNOLOGIES  
 VISUALIZATION  
 LOGS  
 SEARCH  
 SCIENCE  
 DATABASE  
 LARGE  
 INTERNET  
 COLLECTION  
 SERVER  
 EXAMPLES  
 SE



## 后Hadoop时代的大数据架构

文 / 大数据文摘 编辑 / 董飞 图 / 崔峻琦

提到大数据分析平台，不得不说Hadoop系统，Hadoop到现在也超过10年的历史了，很多东西发生了变化，版本也从0.x进化到目前的2.6版本。我把2012年后定义成后Hadoop平台时代，这不是说不用Hadoop，而是像NoSQL（Not Only SQL）那样，有其他的选型补充。我在知乎上也写过Hadoop的一些入门文章 如何学习Hadoop - 董飞 的回答，为了给大家有个铺垫，简单讲一些相关开源组件。

### 背景篇

**Hadoop:** 开源的数据分析平台，解决了大数据（大到一台计算机无法进行存储，一台计算机无法在要求的时间内进行处理）的可靠存储和处理。适合处理非结构化数据，包括HDFS，MapReduce基本组件。

**HDFS:** 提供了一种跨服务器的弹性数据存储系统。

**MapReduce:** 技术提供了感知数据位置的标准化处理流程：读取数据，对数据进行映射（Map），使用某个键值对数据进行重排，然后对数据进行化简（Reduce）得到最终的输出。

**Amazon Elastic Map Reduce(EMR):** 托管的解决方案，

运行在由Amazon Elastic Compute Cloud（EC2）和Simple Storage Service（S3）组成的网络规模的基础设施之上。如果你需要一次性的或不常见的大数据处理，EMR可能会为你节省开支。但EMR是高度优化成与S3中的数据一起工作，会有较高的延时。

Hadoop 还包含了一系列技术的扩展系统，这些技术主要包括了Sqoop、Flume、Hive、Pig、Mahout、Datafu和HUE等。

**Pig:** 分析大数据集的一个平台，该平台由一种表达数据分析程序的高级语言和对这些程序进行评估的基础设施一起组成。

**Hive:** 用于Hadoop的一个数据仓库系统，它提供了类似于SQL的查询语言，通过使用该语言，可以方便地进行数据汇总，特定查询以及分析。

**Hbase:** 一种分布的、可伸缩的、大数据储存库，支持随机、实时读/写访问。

**Sqoop:** 为高效传输批量数据而设计的一种工具，其用于Apache Hadoop和结构化数据储存库如关系数据库之间的数据传输。

Flume: 一种分布式的、可靠的、可用的服务, 其用于高效地搜集、汇总、移动大量日志数据。

ZooKeeper: 一种集中服务, 其用于维护配置信息, 命名, 提供分布式同步, 以及提供分组服务。

Cloudera: 最成型的Hadoop发行版本, 拥有最多的部署案例。提供强大的部署、管理和监控工具。开发并贡献了可实时处理大数据的Impala项目。

Hortonworks: 使用了100%开源Apache Hadoop提供商。开发了很多增强特性并提交至核心主干, 这使得Hadoop能够在包括Windows Server和Azure在内平台上本地运行。

MapR: 获取更好的性能和易用性而支持本地Unix文件系统而不是HDFS。提供诸如快照、镜像或有状态的故障恢复等高可用性特性。领导着Apache Drill项目, 是Google的Dremel的开源实现, 目的是执行类似SQL的查询以提供实时处理。

原理篇

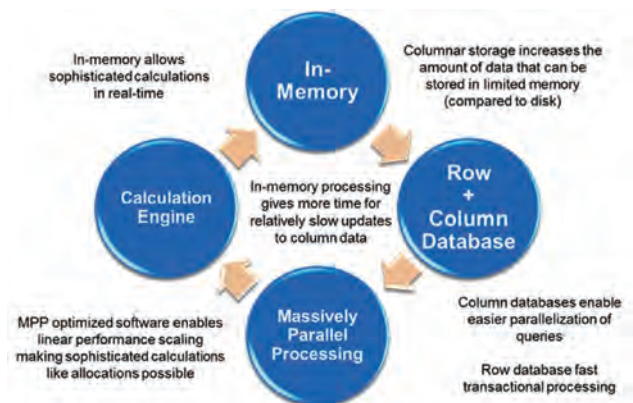
数据存储

我们的目标是做一个可靠的, 支持大规模扩展和容易维护的系统。计算机里面有个locality (局部性定律), 如图所示。从下到上访问速度越来越快, 但存储代价更大。



图 1.9 一个存储器层次模型的示例

相对内存, 磁盘和SSD就需要考虑数据的摆放, 因为性能会差异很大。



磁盘好处是持久化, 单位成本便宜, 容易备份。但随着内存便宜, 很多数据集可以考虑直接放入内存并分布到各机器上, 有些基于 key-value, Memcached用在缓存上。

内存的持久化可以通过 (带电池的RAM), 提前写入日志再定期做Snapshot或者在其他机器内存中复制。当重启时需要从磁盘或网络载入之前状态。其实写入磁盘就用在追加日志上面, 读的话就直接从内存。像VoltDB, MemSQL, RAMCloud 关系型又基于内存数据库, 可以提供高性能, 解决之前磁盘管理的麻烦。

HyperLogLog & Bloom Filter & CountMin Sketch

都是应用于大数据的算法, 大致思路是用一组相互独立的哈希函数依次处理输入。HyperLogLog 用来计算一个很大集合的基数 (即合理总共有多少不相同的元素), 对哈希值分块计数: 对高位统计有多少连续的0; 用低位的值当做数据块。BloomFilter, 在预处理阶段对输入算出所有哈希函数的值并做出标记。当查找一个特定的输入是否出现过, 只需查找这一系列的哈希函数对应值上有没有标记。对于BloomFilter, 可能有False Positive, 但不可能有False Negative。BloomFilter可看做查找一个数据有或者没有的数据结构 (数据的频率是否大于1)。CountMin Sketch在BloomFilter的基础上更进一步, 它可用来估算某一个输入的频率 (不局限于大于1)。

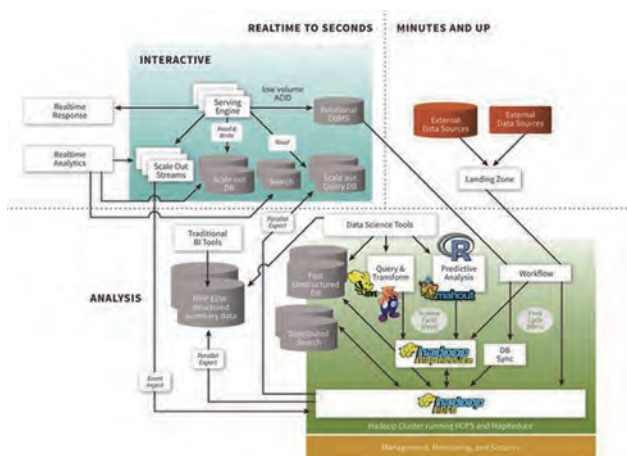
CAP Theorem



简单说是三个特性: 一致性, 可用性和网络分区, 最多只能取其二。设计不同类型系统要多去权衡。分布式系统还有很多算法和高深理论, 比如: Paxos算法 (paxos分布式一致性算法--讲述诸葛亮的反穿越), Gossip协议 (Cassandra学习笔记之Gossip协议), Quorum (分布式系统), 时间逻辑, 向量时钟 (一致性算法之四: 时间戳和向量图), 拜占庭将军问题, 二阶段提交等, 需要耐心研究。



技术篇



根据不同的延迟要求 (SLA)，数据量存储大小，更新量多少，分析需求，大数据处理的架构也需要做灵活的设计。上图就描述了在不同领域中大数据组件。

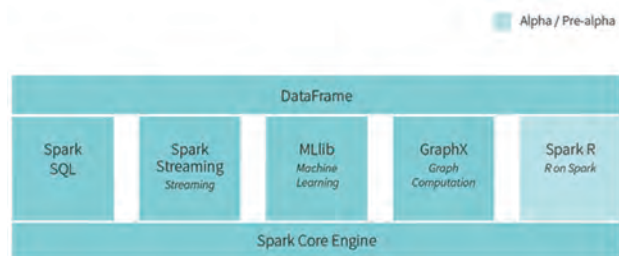
说大数据的技术还是要先提Google，Google 新三辆马车，Spanner, F1, Dremel

Spanner: 高可扩展、多版本、全球分布式外加同步复制特性的谷歌内部数据库，支持外部一致性的分布式事务;设计目标是横跨全球上百个数据中心,覆盖百万台服务器,包含万亿条行记录! (Google就是这么霸气~)

F1: 构建于Spanner之上,在利用Spanner的丰富特性基础之上,还提供分布式SQL、事务一致性的二级索引等功能,在AdWords广告业务上成功代替了之前老旧的手工MySQL Shard 方案。

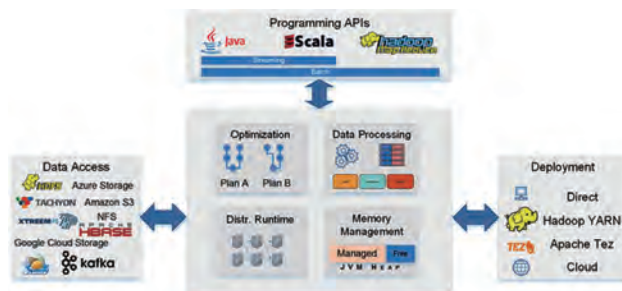
Dremel: 一种用来分析信息的方法，它可以在数千计的服务器上运行，类似使用SQL语言，能以极快的速度处理网络规模的海量数据(PB数量级)，只需几秒钟时间就能完成。

Spark



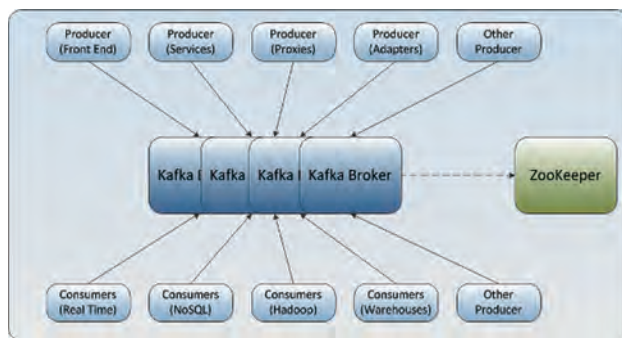
2014年最火的大数据技术Spark，。主要意图是基于内存计算做更快的数据分析。同时支持图计算，流式计算和批处理。Berkeley AMP Lab的核心成员出来成立公司Databricks开发Cloud产品。

Flink



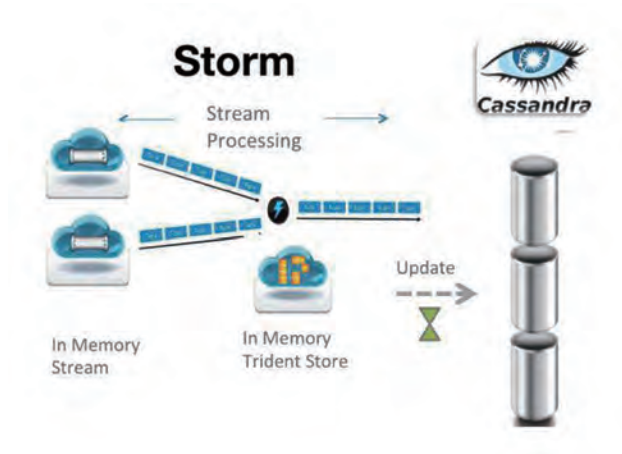
使用了一种类似于SQL数据库查询优化的方法，这也是它与当前版本的Apache Spark的主要区别。它可以将全局优化方案应用于某个查询之上以获得更佳的性能。

Kafka



Announcing the Confluent Platform 1.0 Kafka 描述为LinkedIn 的“中枢神经系统”，管理从各个应用程序汇聚到此的信息流，这些数据经过处理后再被分发到各处。不同于传统的企业信息列队系统，Kafka 是以近乎实时的方式处理流经一个公司的所有数据，目前已经为 LinkedIn, Netflix, Uber 和 Verizon 建立了实时信息处理平台。Kafka 的优势就在于近乎实时性。

Storm



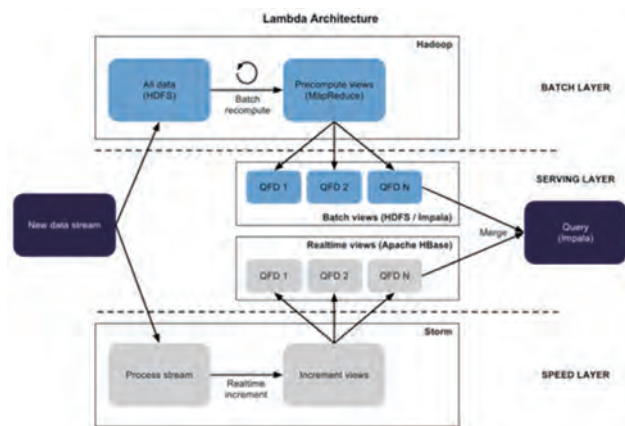
Handle Five Billion Sessions a Day in Real Time, Twitter的实时计算框架。所谓流处理框架，就是一种分布式、高容错的实时计算系统。Storm令持续不断的流计算变得容易。经常用于在实时分析、在线机器学习、持续计算、分布式远程调用和ETL等领域。

Samza

	Storm	Spark	Samza
<b>Delivery Semantics</b>	At Least Once Exactly-Once with Trident	Exactly Once Except in some failure scenarios	At Least Once
<b>State Management</b>	Stateless Roll your own or use Trident	Stateful Writes state to storage	Stateful Embedded key-value store
<b>Latency</b>	Sub-Second	Seconds Depending on batch size	Sub-Second
<b>Language Support</b>	Any JVM-languages, Ruby, Python, Javascript, Perl.	Scala, Java, Python	Scala, Java JVM-languages only

LinkedIn主推的流式计算框架。与其他类似的Spark，Storm做了几个比较。跟Kafka集成良好，作为主要的存储节点和中介。

Lambda architecture



Nathan写了文章《如何去打败CAP理论》How to beat the CAP theorem, 提出Lambda Architecture, 主要思想是对一些延迟高但数据量大的还是采用批处理架构, 但对于即时性实时数据使用流式处理框架, 然后在之上搭建一个服务层去合并两边的数据流, 这种系统能够平衡实时的高效和批处理的Scale, 看了觉得脑洞大开, 确实很有效, 被很多公司采用在生产系统中。

Summingbird

Lambda架构的问题要维护两套系统, Twitter开发了

Summingbird来做到一次编程, 多处运行。



将批处理和流处理无缝连接, 通过整合批处理与流处理来减少它们之间的转换开销。下图就解释了系统运行时。

NoSQL

Document Database	Graph Databases
Couchbase	Neo4j
MarkLogic	InfiniteGraph
mongoDB	
Wide Column Stores	Key-Value Databases
redis	ACCUMULO
amazon DynamoDB	HYPERTABLE™
riak	Cassandra
	APACHE HBASE
	Amazon SimpleDB

数据传统上是用树形结构存储(层次结构), 但很难表示多对多的关系, 关系型数据库就是解决这个难题, 最近几年发现关系型数据库也不行了, 新型NoSQL出现如Cassandra, MongoDB, Couchbase。NoSQL 里面也分成这几类, 文档型, 图运算型, 列存储, key-value型, 不同系统解决不同问题。没一个one-size-fits-all的方案。

Cassandra

大数据架构中, Cassandra的主要作用就是存储结构化数据。DataStax的Cassandra是一种面向列的数据库, 它通过分布式架构提供高可用性及耐用性的服务。它实现了超大规模的集群, 并提供一种称作“最终一致性”的一致性类型, 这意味着在任何时刻, 在不同服务器中的相同数据库条目可以有不同的值。

SQL on Hadoop

开源社区业出现了很多 SQL-on-Hadoop的项目, 着眼



跟一些商业的数据仓库系统竞争。包括Apache Hive, Spark SQL, Cloudera Impala, Hortonworks Stinger, Facebook Presto, Apache Tajo, Apache Drill。有些是基于Google Dremel设计。

Impala

Cloudera公司主导开发的新型查询系统，它提供SQL语义，能够查询存储在Hadoop的HDFS和HBase中的PB级大数据，号称比Hive快5-10倍，但最近被Spark的风头给罩住了，大家还是更倾向于后者。

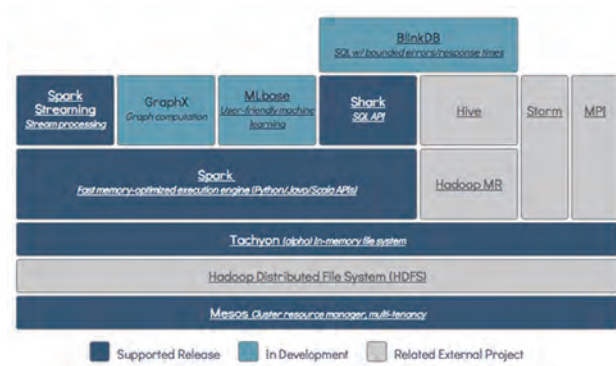
Drill

Apache社区类似于Dremel的开源版本—Drill。一个专为互动分析大型数据集的分布式系统。

Druid

在大数据集之上做实时统计分析而设计的开源数据存储。这个系统集成了一个面向列存储的层，一个分布式、shared-nothing的架构，和一个高级的索引结构，来达成在秒级以内对十亿行级别的表进行任意的探索分析。

Berkeley Data Analytics Stack



上面说道Spark，在Berkeley AMP lab 中有个更宏伟的蓝图，就是BDAS，里面有很多明星项目，除了Spark，还包括：

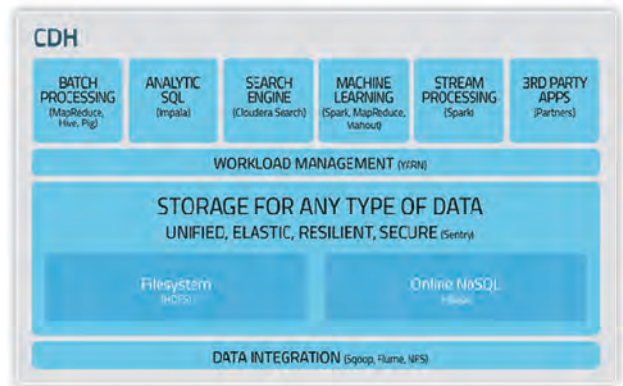
Mesos：一个分布式环境的资源管理平台，它使得Hadoop、MPI、Spark作业在统一资源管理环境下执行。它对Hadoop2.0支持很好。Twitter，Coursera都在使用。

Tachyon：是一个高容错的分布式文件系统，允许文件以内存的速度在集群框架中进行可靠的共享，就像Spark和MapReduce那样。项目发起人李浩源说目前发展非常快，甚至比Spark当时还要惊人，已经成立创业公司Tachyon Nexus。

BlinkDB：也很有意思，在海量数据上运行交互式 SQL 查询的大规模并行查询引擎。它允许用户通过权衡数据精度来提升查询响应时间，其数据的精度被控制在允许的误差范围内。

Cloudera

Hadoop老大哥提出的经典解决方案。



HDP (Hadoop Data Platform)



Hortonworks 提出的架构选型。

Redshift

Amazon Redshift架构

- 主节点
  - SQL语言接口
  - 存储元数据
  - 协调查询执行
- 计算节点
  - 本地，列存储
  - 并行执行查询
  - 通过Amazon S3来加载、备份、恢复；从Amazon DynamoDB或SSH加载
- 两个硬件平台
  - 优化数据处理
  - DW1: HDD; 2TB to 1.6PB
  - DW2: SSD; 160GB to 256TB



Amazon RedShift是 ParAccel一个版本。它是一种 (massively parallel computer) 架构，是非常方便的数据仓库解决方案，SQL接口，跟各个云服务无缝连接，最大特点就是快，在TB到PB级别非常好的性能，我在工作中也是直接使用，它还支持不同的硬件平台，如果想速度更快，可以使用SSD。FIN

# 零售企业库存管理分析

文 / 中颢润项目数据分析师事务所 编辑 / 张楠

随着零售行业竞争的不断加剧，零售企业越来越关注企业的库存水平。降低库存，增加现金流，可以间接提高企业的利润率。那么，如何通过实用、高效的库存管理技术来建立竞争优势并赢得顾客是企业管理的重心。针对零售企业的库存管理我们给出以下管理模式：

## 一、传统库存管理模式

关于库存管理的理论很多，相对于新兴的VMI、JMI、CPER等理论，有些早期的理论，中颢润认为，主要包括了ABC管理法、经济批量模式以及定期和定量订货方法等等。在早期的库存管理中，都是企业独立管理自己的库存，以自身成本的最小化为目标确定库存管理策略，明确订货点、订货周期及数量等等。在这种模式下，生产企业和零售企业的信息都是孤立的，为满足市场都必须承担比较大的库存水平，从而形成“牛鞭效应”。这种模式下，协作缺乏，社会资源效率低下。

### (1)、ABC分类法

对于库存的管理应当有着重点，根据他们的重要性给予不同的关注，将各种资源按重要性进行分配，这种分配方法我们称为ABC分类管理法。就如它的名字，将库存按照重要性的大小划分为A类、B类和C类，然后将资金以及人力和物力按照重要性分类分配给库存。根据ABC管理法的实施经验总结，可以得到一个参考的分类标准，如表1所示：

等级	消耗额(%)	种类数
A	60-80	10-20
B	15-40	20-30
C	5-15	50-70

表1 ABC分类法的消耗额参考标准

### (2)经济批量法

经济批量模式是以成本最小化为目标，在需求已知的情况下，求得订货批量的最优值。在库存管理成本中，包括了订货费用和储存费用，二者都与批量有关。如果批量较小，那么在需求一定的条件下，进货次数就会增加，因而订货费用就会上升，但是同时批量较小则库存水平就会较低，从而使得储存费用下降。

因此，这个模型是在这相互矛盾的二者中寻求一个最佳的平衡点，将库存成本最小化，而这一理论的具体模型可以表示为如下形式：

$$Q^* = \sqrt{\frac{2DS}{C}}$$

其中Q\*——经济批量

D——年需求量

S——订货成本/次

C——储存费用/单位产品

### (3)定量订货法

定量订货法借鉴了经济批量模型，在库存水平减少到一定的程度时，就按照一定的数量订货，而每次进货数量是相等的。

### (4)定期订货法

定期订货法主要与定量订货相对，它是提前确定一个订货周期，每隔这一周期就进行订货作业，但每次订货的批量一般不同。

## 二、合作库存管理模式

随着资本力量的壮大，部分生产企业与零售企业相互控股，协作发展。像泰国正大集团与易初莲花购物中心的合作，易初莲花是正大集团的子公司，在易初莲花的大卖场里，你可以看到正大集团的产品，像肉类、油等，被陈列在最好的堆头和货架上，销售利益双方分享。中颢润认为在这种模式下，双方的信息可以实现充分的共享，这对于生产、销售计划以及促销活动的执行都是非常有利的。

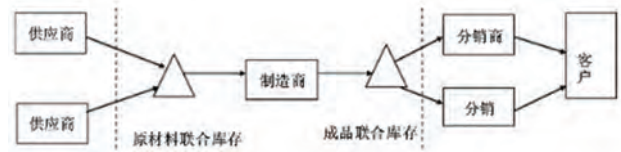


图1 合作库存管理

## 三、供应商维护库存模式

中颢润认为，目前，很多大型商业企业都在逐步实行由供应商管理、维护卖场的库存模式，在这种模式下，零售商负



责订货,并对货物拥有所有权,货物款项由零售商与供应商约定在一定的时间定期结算。供应商派促销员到零售企业的卖场负责本公司产品的库存维护以及促销的工作,促销员负责对产品盘点,并处理退换货,反馈市场信息等功能,促销员作为零售商与供应商的协调员。

在这种库存控制策略下,适当允许上游生产商对下游零售商的库存策略、订货策略进行了解和沟通的权利,双方可以

协同进行促销活动,以增进销售。最重要的是,生产企业可以获得零售商的销售信息部分共享,能够比较全面的预计市场状况,对于零售企业的生产计划安排非常有利。FIN

## COX回归模型在电信用户离网分析中的应用

文 / 河南智宸项目数据分析事务所 宁威 编辑 / 李子昂

随着国资委对三大运营商营销费用的削减以及虚拟运营商开始逐渐发力,电信行业的竞争日益加剧。为了加强市场竞争力,各大电信运营商纷纷构建了自己的数据仓库系统,将用户的各种相关数据进行集中和整理,然后从各个角度分析用户的行为,从而为确定具体的营销活动提供参考依据。建立在分析基础上的营销活动,不但成本较低,而且极大地提高了营销活动的成功率。

用户离网分析是建立数据仓库系统后第一个主要的数据挖掘应用。本文基于某市电信运营商数据库获取的数据,进行用户的流失分析,采用生存分析中的cox回归模型,通过spss statistics21分析用户各项基础信息及行为数据,得出导致用户流失的因素及时间规律,为企业建立用户离网预警机制提供了有力的数据支撑。

### 用户离网分析步骤

CRISP-DM数据分析建模指导广泛用于数据的挖掘,是一种数据挖掘的生命周期模型。

#### (1)问题定义

模型中,我们主要关心的是哪些因素造成用户流失,这些因素在哪些时间提高了用户的生存率和危险率。

#### (2)数据准备

数据准备阶段,模型中收集了主要包括了用户的基本信息、用户消费信息、用户行为信息(业务信息)。

#### (3)建模

建模是根据相关的算法,产生评估用户离网概率模型的过程,包括:1.根据目标变量的分布特征确定抽样策略,生成目标数据集;2.数据描述/探索过程,了解数据的分布以及数据之间的关系;3.调整属性集,增加、删除或衍生用户(用户)

的属性,并针对不同算法进行不同的数据预处理;4.通过数据分割生成最终训练数据集,并校验和测试数据集。

### 概念介绍

死亡概率:期初变量在单位时间段内死亡的概率,表示为:  $q$

生存概率:单位时间段内,生存个体到结束存活的概率,表示为:  $1 - q$

生存函数:生存率随时间的变化而变化,表示为:  $S(t)$

Cox模型又称比例风险模型,分析发生某种结局(如死亡)的时间与危险因素的联系。Cox模型可理解为两个部分:基础风险(发生结局事件)函数,常用  $\Lambda_0(t)$ 表示,表示随时间变化的发生结局事件的风险;效应参数,表示当某种暴露(危险)因素存在时,基础风险函数的变化。基本结构为:  $h(t) = h_0(t) \cdot \exp(b_1 X_1 + b_2 X_2 + \dots + b_n X_n)$ 。上式中  $X, b$  为影响预后的因素及其回归系数,模型中  $h_0(t)$  是基准风险函数,表示当所有因素(自变量)  $X$  都取0时的危险度。 $h_0(t)$  随时间而变化,其函数类型不作限定,这样就避免了对时间变量的分布作约束或假定,一般不能由样本估计出。 $\exp(b_n X_n)$  表示第  $n$  个因素  $X$  所产生的作用,此项作用就是使死亡相对风险由  $h_0(t)$  增至  $h_0(t) \cdot \exp(b_n X_n), n = 1, 2, 3, \dots, n$ 。

Cox回归方程是由各危险因素的Cox回归系数和标准化Cox回归系数(回归系数与标准误之比),反映和比较各危险因素对某种危险函数或相对危险度的作用。预报是根据一个用户的各危险因素取值,来预测用户从某种起始事件开始的未来各时刻的生存概率。

### 数据处理

数据简单性描述:

通过上述简单性描述，从热图中可以看出：服务月数长的用户全体主要集中在2、3、4类别的用户；

在选定变量后，我们将用户流失作为目标变量，在网月长作为时间段，将剩下的变量指标选入协变量框中，在分类中，选择家庭人数、用户类别、性别、是否退休作为分类协变量（作为指示符），参考类别为：最后一个；在绘图中我选择生存函数、危险函数，将用户类别作为协变量值的位置；保存，选择函数、危险函数、标准误差；选项，选择CI、估计相关性。用单因素Lo-rank检验方法进行初步筛选，得出11个因素对用户流失有显著影响，分别是：地址、工作年数、设备租赁、电话卡服务、上月长途、上月设备、多网路、语音邮箱、呼叫转接、三方呼叫、电子账单（ $p < 0.05$ ）

结果分析

通过分析可以得到这11个因素的回归方程为：

$$h(t) = h_0(t) * \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11})$$

- a. Variable(s) Entered at Step Number 1: 上月长途
- b. Variable(s) Entered at Step Number 2: 设备租赁
- c. Variable(s) Entered at Step Number 3: 工作年数
- d. Variable(s) Entered at Step Number 4: 多网路
- e. Variable(s) Entered at Step Number 5: 电子账单
- f. Variable(s) Entered at Step Number 6: 电话卡服务
- g. Variable(s) Entered at Step Number 7: 语音邮箱
- h. Variable(s) Entered at Step Number 8: 上月设备
- i. Variable(s) Entered at Step Number 9: 地址
- j. Variable(s) Entered at Step Number 10: 呼叫转接
- k. Variable(s) Entered at Step Number 11: 三方呼叫
- l. Beginning Block Number 1. Method = Forward Stepwise (Likelihood Ratio)

由定义可知，风险函数为：

$$\lambda = \exp(-0.037 X_1 - 0.051 X_2 + 1.879 X_3 - 0.072 X_4 - 0.232 X_5 - 0.04 X_6 - 0.759 X_7 + 0.576 X_8 + 0.698 X_9 - 0.45 X_{10} + 0.77 X_{11})$$

通过分析可以得到这11个因素的回归方程为：

$$h(t) = h_0(t) * \exp(-0.037 X_1 - 0.051 X_2 + 1.879 X_3 - 0.072 X_4 - 0.232 X_5 - 0.04 X_6 - 0.759 X_7 + 0.576 X_8 + 0.698 X_9 - 0.45 X_{10} + 0.77 X_{11})$$

图中说明协变量：服务月数在不同水平下，不同用户类型的死亡率分布中，当服务月数小于40时，四种用户类型的死亡率分布差异不大，而当服务月数大于40时，用户类型中死亡率的分布开始有明显差异，分布情况为：电子服务>附加服务>总服务>基础服务的分布规律。

分析结论

通过上述分析得出的结果，可以发现影响用户生存时间的因素有：地址、工作年数、设备租赁、电话卡服务、上月长途、上月设备、多网路、语音邮箱、呼叫转接、三方呼叫、电子账单11个因素（ $p < 0.05$ ）。当服务月数即将达到40时，应进行针对性的通过电子服务、附加服务的营销活动，提前加大对“待”流失客户的维系工作。

讨论

Cox回归模型的优越性之一就是对基准生存分布没有特殊要求，所以称为半参数回归模型，在实用方面比参数回归模型更为灵活，能够迅速方便地提供影响预后的参考因素，以利于实际应用中更早地发现潜在的流失用户，提高生存率或延长用户使用时间。

本例中1000例用户“预后”的生存分析中，Cox回归模型提供了主要的“预后”影响因素，针对用户流失研究提供早发现、早准备和早营销的有力理论依据。由于本案要求时间短，以及数据完整性等问题，暂未彻底地评估模型，检查构造模型的步骤，同时还需要确定是否有重要业务问题没有被充分的考虑。有一些结论也还有待于进一步观察，特别是对那些影响用户生存率的重要因素，有待于今后深入的探讨。FIN



## 河南智宸项目数据分析师事务所

河南智宸项目数据分析师事务所有限公司（以下简称“智宸事务所”）经中国商业联合会数据分析专业委员会严格考察批准已正式入会（团证105号）。

智宸事务所现已汇集通信、零售、交通运输等各方精英，并聘请众多数据分析行业专家及高校教授作为资深顾问团队。团队核心人员均为研究生以上学历，拥有多家大型企业数据分析行业经验。

智宸事务所将以可靠的数据来源、专业的分析技术、公正的原则、诚信的精神，为河南的广大企业（主要方向为通信业、零售业、交通运输业）、银行金融机构、政府组织机构提供投资项目数据分析（评估、分析、规划、策划），企业经营数据分析，企业战略分析等具有经济性、权威性、客观性、公正性、实用性的项目数据分析报告及其他行业咨询类报告。

目前事务所已与多家企业、政府、高校建立了合作关系，将利用自身人才及行业优势共同推进河南数据分析行业发展。同时已加入河南MBA&EMBA企业家联合会，作为数据分析顾问单位服务于广大会员企业。

智宸的理念：精准的数据视角+专业的行业经验+优质的全程服务



智宸的使命：致力于成为中部地区最大、最专业的数据分析机构，为中部地区政府及企业发展提供独立、权威、公正、诚信、优质的“数据分析”服务。同时肩负推广河南地区数据分析行业发展的重任。

联系方式：孙经理 0371-55339936 / 17737707797

办公地址：郑州市农业路东16号省汇中心B座2408室

网站主页：www.hnzcdata.com

企业邮箱：Info@hnzcdata.com

## 山东海洋项目数据分析师事务所

山东省海洋项目数据分析师事务所，重点致力于数据分析、决策统筹、系统工程的研究与应用，坚持“数据分析、信息挖掘、决策体系”三结合原则，密切追踪各行业发展趋势，聚焦国内产业政策及公司运作与治理研究，保持与企业、投



行、政府职能部门、研究机构、协会的良好合作关系，秉承“独立、客观、公正、高端”的执业理念，专业、高效开展项目数据分析（包括但不限于：行业数据分析、经营数据分析、投资数据分析、运营风险数据分析、决策数据分析），战略咨询，投融资设计与服务，资本运作咨询等综合业务，为客户决策持续提供最具价值的服务。

山东海洋CPDA，重点着眼并服务于三大决策效能的提升：运营决策、管理决策、投融资决策。

山东海洋CPDA，值得您信赖的决策参谋！

联系方式：柴经理 15106949297

办公地址：济南市经十路20637号新华文博楼4F-C区

网站主页：www.sdhycpda.com

企业邮箱：sdhycpda@sina.com



# 智宸数据

值得信赖的数据分析专家

- 智慧决策从数据开始
- 用数据说话，做理性分析
- 通过深度分析为客户解决经营难题
- 创造企业数据核心价值

## 河南智宸项目数据分析师事务所简介

河南智宸项目数据分析师事务所是河南省最为专业的数据分析机构。通过大数据分析 & 企业深度剖析建立精准模型，为企业规避经营风险、优化商业决策提供大数据解决方案，同时为政府构建智慧城市提供量化数据支持。

**智宸数据的理念：**精准的数据视角+专业的行业经验+优质的全程服务

**智宸数据的使命：**致力于成为中部地区最大、最专业的数据分析机构，为中部地区政府及企业发展提供独立、权威、公正、诚信、优质的“数据分析”服务。

**0371-55339936**

**河南智宸项目数据分析师事务所有限公司**  
HENAN ZONE CHANCE PROJECT DATA ANALYST FIRM LIMITED

Add:河南省郑州市农业路省汇中心B座2408( 450008 ) Fax:0371-85969996 Email:hncdata@163.com Web:www.hncdata.com