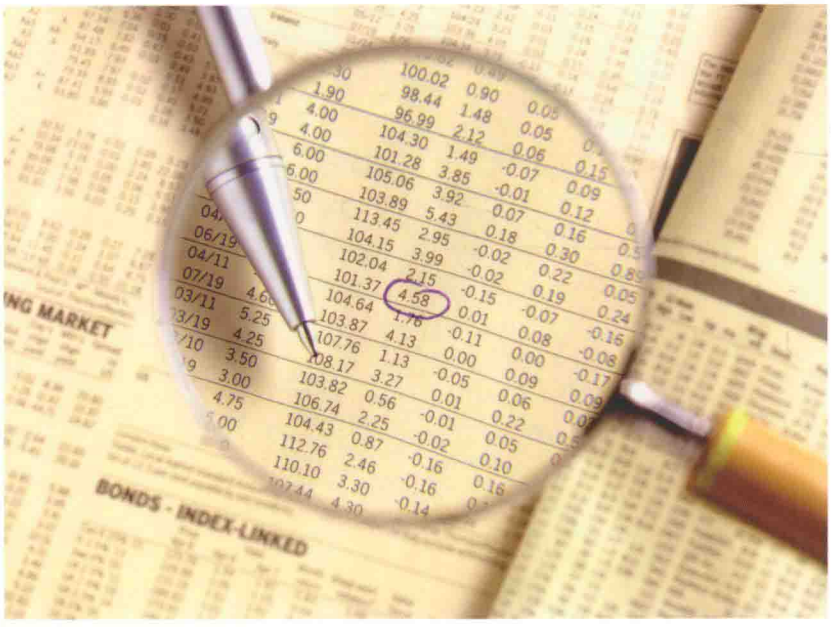
数据分析员（CDA）考试丛书

CDA数据分析

零基础入门

中国商业联合会数据分析专业委员会　编著



中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

数据分析员(CDA)考试丛

CDA数据分析——零基础入门

中国商业联合会数据分析专业委员会　编著

電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

内 容 简 介

该书基于通用的 Excel、SPSS 工具，加上必知必会的数据分析概念，以图文并茂、理论与实操相结合的方式，按照 CDA 人才培养考核要求进行编写。全书分为 6 章，分别为数据分析概述、数据收集与导入、数据的清洗与预处理、数据可视化呈现、基础数据分析、综合分析。

本书适合数据分析零基础群体读者阅读，也可供大学生、初入数据分析职场人员、参与 CDA 考试的人员学习使用。

图书在版编目(CIP)数据

CDA 数据分析：零基础入门/中国商业联合会数据分析专业委员会编著. —北京：电子工业出版社，2016.4
ISBN 978-7-121-28475-5

Ⅰ.①C… Ⅱ.①中… Ⅲ.①商业统计-数据处理 Ⅳ.①F712.3

中国版本图书馆 CIP 数据核字(2016)第 060932 号

策划编辑：石会敏
责任编辑：王二华
印　　刷：三河市双峰印刷装订有限公司
装　　订：三河市双峰印刷装订有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　　邮编：100036
开　　本：787×1092　1/16　印张：12.75　字数：300 千字
版　　次：2016 年 4 月第 1 版
印　　次：2016 年 4 月第 1 次印刷
印　　数：4000 册　　定价：38.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010)88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：(010)88254537，88254532。

写给读者

全球大数据时代到来了，不仅因为IT技术的变革使大数据得以产生和膨胀，因为大数据分析使原来不可能的精确决策成为可能，因为国家将大数据提升为国家战略，因为中国专业的数据分析师、数据分析师事务所正在崛起，更深层的原因在于：人们探究科学、探究真理的努力从未停息。大数据将改变我们所有人、所有企业的行为轨迹和思维方式。

IDC咨询公司对2020年大数据市场的一组预测显示：到2020年，65%的大型企业会将自己武装成数据化公司；大数据分析市场将以23%的年复合增长率高速发展；利用大数据分析技术，各行业将在生产力方面节省超过4000亿美元；未来的全球2000强公司中将有85%的企业是借助数据化转型之机树立行业领先地位的……

大数据时代，无论大家是否介意大数据可能产生的问题（如个人隐私数据的泄露、人工智能带来的风险等），大数据对人们的改变都无法避免，未来的世界是数据化世界，未来的企业一定是数据化企业，业务依靠数据而延展，人们依靠数据的分析能力来证明自己的实力。美国大数据人才需求分析报告显示：到2018年，美国数据分析师的人才需求将达150万人左右。我国相关部门统计，预计未来3～5年内，我国数据分析专业人才的需求将达100万人以上。

到目前为止，在我国所有的高校中，还没有真正意义上的数据分析专业，因为数据分析不仅需要数学、统计分析、信息管理等知识，更重要的是要把数据决策的思维与企业的运营思维相结合，“做分析”比“做数据”更有意义、更重要。数据分析为企业带来的不是“概念”和“故事”，而是真正通过精确决策帮助企业实实在在地省钱和赚钱，这样的数据分析人才，在今天和未来的商业价值都无可估量。

当然要成为真正的数据分析人才，不仅要掌握多元化的知识体系，更要将知识与实践结合。现在获取数据的方式和途径越来越丰富，如政府或行业的公开数据、互联网爬虫收集等，运作数据分析的方法和工具，对数据进行处理，然后分析、分析、再分析，尽力探寻数据背后的规律是每一个从事数据分析工作的个人进入大数据殿堂的必由之路。

中国商业联合会数据分析专业委员会作为我国大数据分析领域的行业带头人，深知数据分析人才的价值和重要，我们积极推动数据分析员（CDA）、数据分析师（CPDA）的培养体系的建立，希望通过本书帮助高校的大学生们提高对数据的兴趣、培养数据分析的素质，也希望由此能促进我国高校数据人才培养工作的开展。

面对越来越激烈的就业市场，选择一个朝阳产业、一个金领职业，是每个将要进入社会的大学生需要谨慎思考的事情。如果将优秀职业的属性数据进行分析，你会发现数据分析的标签精彩而动人：企业对数据分析人才的需求将越来越旺盛、数据分析可以游刃有余地适应各个行业、薪酬诱人、数据分析经验日久弥新……

所以，如果你想成为我们中的一员，那就从现在开始拥抱数据吧！

中国商业联合会数据分析专业委员会会长　邹东生

2016 年 3 月

前　言

随着大数据概念的推广与普及，数据正在像石油、钢铁一样成为重要的原材料，以数据为重要驱动力的数据革命正在到来。相应人才能力的培养重点也在变革，尤其是整合企业数据的能力、探索数据背后价值和制定精确行动纲领的能力、进行精确快速实时行动的能力。

在数据化时代，人们将以各种数据为工作对象，将数据与传统产业结合起来，为帮助读者快速具备科学的数据分析思维，提升数据分析能力，本书在内容设计上满足了广大数据分析初学者渴望全面学习数据分析的要求。我们编写本套丛书，希望能够让学习者掌握数据分析思维能力，将技能运用到企业需要的岗位中，将能力转化为真正的价值。

本套丛书是在中国商业联合会数据分析专业委员会考试专家的指导下编写完成的。在编写过程中根据数据分析初学者的学习习惯，采用由浅入深、由易到难的方式讲解，读者还可以通过随书赠送的多媒体视频教学课程学习。本套丛书结构清晰，内容丰富，主要包括以下三册。

◆《CDA 数据分析考试大纲》

本大纲是全国数据分析员职业技能水平考试的标准和命题依据，是专业技术人员能力测评和指导专业学习的依据。本大纲包括《CDA 数据分析——零基础入门》、《CDA 数据分析实务》两科考试的内容和范围，即数据分析思维能力考核，基础数据分析技术，数据采集、清洗、加工整理和图标展示等技术展现，是理论性、技术性和实践性很好的结合。

◆《CDA 数据分析——零基础入门》

本书从理论层面解读大数据思维能力的培养，详解大数据基础能力培养的步骤，透过案例讲知识。教材中，概念、原理及理论叙述准确、精炼，知识点突出，难点分散，算法过程严谨，具有代表性和启发性，适应普通高等学校层次教学的需要。

◆《CDA 数据分析实务》

本书侧重在企业实际经营过程中数据价值的发挥，针对企业中不同业务部门的活动、不同业务决策所需要的数据分析，提供了各种模型和算法的运用。

三本书是一个相对完整的体系，各有侧重。总结起来，本套丛书主要有以下特点。

1. 将数据分析方法和实务操作相结合，突出该学科的方法论作用。

2. 针对数据分析业务活动的实用性和操作性的特点，理论、操作和实务相结合，有利于读者全面掌握理论和应用。

3. 本书提供了丰富的全真案例。在实践部分提供的真实资料基础上，本书精选若干典型案例，为读者提供了比较全面的数据分析经验。

本套丛书为全国数据分析员专业技术考试指定教材，也可作为财政、金融、投资咨询等行业的企业经营分析、管理人员的数据分析方法学习用书或工作中的参考书。

本套丛书由中国商业联合会数据分析专业委员会(CDAC)主持编写，中国工信出版集团电子工业出版社负责出版。除主要编写人员外，还有很多专家也为本套丛书的编写和出版工作提供了宝贵的建议和意见，在此对他们的辛勤工作表示衷心的感谢！在本套丛书的编写工作中得到了工业和信息化部教育与考试中心的大力支持和帮助，在此表示特别的感谢！我们还要感谢中国工信出版集团电子工业出版社的编辑，正是他们的认真工作才使本书顺利出版。

由于书中概念和术语数目繁多，书中有不当之处，恳请读者批评指正。我们的电子邮箱：services@chinacpda.org。

中国商业联合会数据分析专业委员会教材编写专家组

2016 年 3 月

目　录

第 1 章

数据分析概述

1.1 数据分析行业发展

数据是事实，也称观测值，是实验、测量、观察、调查等的结果，常以数量的形式给出。数据分析(Analysis of Data)是组织有目的地收集数据、分析数据，使之成为信息的过程。

1.1.1 大数据行业背景和发展趋势

一、大数据与数据分析

1. 大数据

随着大数据概念的普及，人们常常会问，多大的数据才叫大数据？其实，关于大数据，难以有一个非常定量的定义。维基百科给出了一个定性的描述：大数据是指无法使用传统和常用的软件技术和工具在一定时间内完成获取、管理和处理的数据集。进一步来说，当今“大数据”一词的重点其实已经不仅在于数据规模的定义，更代表着信息技术发展进入了一个新的时代，代表着爆炸性的数据信息给传统的计算技术和信息技术带来的技术挑战和困难，代表着大数据处理所需的新的技术和方法，也代表着大数据分析和应用所带来的新发明、新服务和新的发展机遇。

大数据作为时下火热的词汇，随之而来的数据仓库、数据安全、数据分析挖掘等围绕大数据商业价值的利用逐渐成为行业人士争相追捧的利润焦点。随着大数据时代的来临，大数据分析也应运而生。早在2010年12月，美国总统办公室下属的科学技术顾问委员会(PCAST)和信息技术顾问委员会(PITAC)就向奥巴马和国会提交了一份《规划数字化未来》的战略报告，把大数据收集和使用的工作提升到体现国家意志的战略高度。报告列举了5个贯穿各个科技领域的共同挑战，而第一个最重大的挑战就是“数据”问题。报告指出：“如何收集、保存、管理、分析、共享正在呈指数增长的数据是我们必须面对的一个重要挑战。”

大数据最大的特点：它是针对传统手段捕捉到的数据之外的非结构化数据。这意味着不能保证输入的数据是完整的、清洗过的和没有任何错误的。这就使它更有挑战性，且同时提供了在数据中获得更多的洞察力的范围。在典型的世界里，很难在所有的信息间以一种正式的方式建立关系，因此非结构化数据以图片、视频、移动产生的信息、无线射频识别(RFID)等的形式存在，这些都被考虑进大数据分析的范畴。绝大多数的大数据分析数据库基于纵列数据库之外。大数据分析是利用对数据有意义的软件支持针对于数据的实时分析。当市场上有大数据分析的应用系统时，同样可以通过通用的硬件和新一代的分析软件，像 Hadoop 或其他分析数据库来实现。

不断积累的大数据包含着很多在小数据量时不具备的深度知识和价值，大数据分析将为行业/企业带来巨大的商业价值，实现各种高附加值的增值服务，进一步提升行业/企业的经济效益和社会效益。

2. 数据分析

数据分析是用包括检查、清洗、转换和建模等方法对数据进行处理。其目的是探索有用的信息、给出有建设性的意见和辅助制定决策。数据分析包含很多方面和方法，涉及的领域也遍布经济、科学、社会福利等行业。

数据挖掘是一个特别的数据分析技术，与传统的以纯描述为目的的技术相比，更专注于预测模型和潜在知识的挖掘。预测分析用统计模型预测或矫正。商业智能依靠数据分析的深度挖掘和重组聚合挖掘商业信息。数据分析不仅能从真实的数据中发现问题，而且还能通过经济学原理建立数学模型，对投资或其他决策的可行性进行分析，预测未来的收益及风险情况，为科学合理的决策提供依据。在提高工作效率的基础上，也增强了企业管理的科学性。在应用统计里，人们把数据分析分为描述统计、挖掘数据分析和证实数据分析。挖掘数据分析关注于探索新的数据特征，证实数据分析帮助证明已存在假设的真伪。所有的这些都是数据分析。数据分析虽然没有数据科学那样先进的可以创新的数据结构，但是他们的目的是一样的——探索数据可以用来怎样回答问题和解决问题。

3. 大数据 1.0 到大数据 2.0 的发展

由维克托·迈尔·舍恩伯格编写的《大数据时代》里指出大数据是指采用所有数据进行分析，而不是抽样调查。大数据有 4V 特点：Volume(大量)、Velocity(高速)、Variety(多样)、Value(价值)。大数据 1.0 时代的特征是解决数据效率问题，大数据时代 4 个 V 中的前 3 个 V 都被有效地诠释了。但是最后一个 V(Value 价值)还没有表现其作用。因此很多人认为大数据的炒作概念超过了实际价值，也有人认为大数据的概念是美国 IT 巨头为销售其产品的炒作。Hadoop、Hive、Map reduce、R 语言、Python，成为了大数据 1.0 时代的热词。

从理论上看数据信息已经存在很久了，但是数据信息的价值的探讨和研究则是最近几年爆发的。大数据 1.0 时代逐一地解决了速度、容量等问题。在 1.0 时代里搜寻大数据已经变得很方便了。从历史上看，技术的不断突破必然导致大数据时代的不断成熟。从 1.0

进入2.0时代是必然的结果。在1.0时代积累的大数据将会在2.0时代得到其在价值上面的发挥。

从2015年起，大数据进入了2.0时代。大数据2.0时代要求以数据本身的价值为目标，从企业本身业务需求产生的大量数据中通过深入挖掘、分析得出数据本身的价值。1.0停留在数据认知上面，2.0则要求通过这些数据去解决问题。金融行业作为领头羊首先尝试突破传统，积极改革应对大数据2.0时代的冲击。银行作为金融行业的根基已经从传统的ATM取款机、信用卡走出来了，银行要结合大数据营销来获取客户，开发新的理财产品。近期的线上理财、短期理财、网贷等网络理财产品的崛起将为银行带来新的收入，同时也为客户提供全新的理财感受。

二、大数据分析的国际背景

在全球500强企业中，90%以上的重要投资与经营决策都取决于充分的数据分析支持。在欧盟、美国、日本等发达地区，数据分析普遍被作为运营决策的前提要素，为社会经济的高速发展作出了巨大贡献。

美国政府将大数据视为强化美国竞争力的关键因素之一，把大数据研究和生产计划提高到国家战略层面。2012年3月，美国奥巴马政府宣布投资2亿美元启动“大数据研究和发展计划”，这是继1993年美国宣布“信息高速公路”计划后的又一次重大科技发展部署。美国政府认为大数据是“未来的新石油与矿产”，将“大数据研究”上升为国家意志，对未来的科技与经济发展必将带来深远影响。美国政府还在积极推动数据公开，已开放37万个数据集和1209个数据工具，并在2013年5月月初进一步要求，政府必须实现新增和经处理数据的开放和机器可读，激发大数据创新活力。同时，美国政府也是大数据的积极使用者，2013年曝光的“棱镜门”事件显示出美国国家安全部门大数据应用的强大实力，其应用范围之广、水平之高、规模之大都远远超过人们的想象。2012—2013年，美国国家安全局(NSA)、联邦调查局(FBI)及中央情报局(CIA)等联邦政府机构还大量采购亚马逊的云服务，以支撑其大数据应用。

数据分析行业在2012年美国职业调查评选中被评为最性感的行业，当下越来越多的人开始关注这个行业。美国数据分析行业的薪酬情况：虽然薪酬的多少要看个人的分析水平，技术越高越能拿到高的工资，但是对于一般能够掌握基础的数据分析工具的数据分析师来说，每年能拿到35 000~45 000美元。丰厚的工资也使得越来越多的大学生关注这个行业。

继美国率先开启大数据国家战略先河之后，欧盟、日本及韩国等国家也将跟进，预计不久相应的战略举措也将出台。数据规模及运用数据的能力将成为综合国力的重要组成部分，对数据的占有和控制也将成为国家间争夺的焦点。

英国政府紧随美国之后，推出一系列支持大数据发展的举措。首先是给予研发资金支持。2013年1月，英国政府向航天、医药等8类高新技术领域注资6亿英镑支持研发，其中大数据技术获得1.89亿英镑的资金，是获得资金最多的领域。其次是促进政府和公共领

域的大数据应用。据测算，通过合理、高效使用大数据技术，英国政府每年可节省约330亿英镑，相当于英国每人每年节省约500英镑。为了在医疗领域更好地应用大数据，2013年5月，英国政府和李嘉诚基金会联合投资设立全球首个综合运用大数据技术的医药卫生科研机构，将透过高通量生物数据，与业界共同界定药物标靶，处理目前在新药开发过程中关键的瓶颈，之后还将汇集遗传学、流行病学、临床、化学和计算机科学等领域的顶尖人才，集中分析庞大的医疗数据。

法国政府为促进大数据领域的发展，将以培养新兴企业、软件制造商、工程师、信息系统设计师等为目标，开展一系列的投资计划。法国政府在其发布的《数字化路线图》中表示，将大力支持包括“大数据”在内的战略性高新技术，法国软件编辑联盟曾号召政府部门和私人企业共同合作，投入3亿欧元资金用于推动大数据领域的发展。

日本政府认为，提升日本竞争力，大数据应用不可或缺。日本在新一轮IT振兴计划中把发展大数据作为国家战略的重要内容，新的ICT战略重点关注大数据应用技术。日本总务省2012年7月推出了新的综合战略“活力ICT日本”，将重点关注大数据应用，并将其作为2013年六个主要任务之一，聚焦大数据应用所需的、社会化媒体等智能技术开发，以及在新医疗技术开发、缓解交通拥堵等公共领域的应用。

2013年8月月初，澳大利亚出台公共服务大数据政策，提出了大数据分析的实践指南，希望通过大数据分析系统提升公共服务质量，增加服务种类，为公共服务提供更好的政策指导。在新加坡政府，多个国际领先企业在当地设立大数据技术研发中心，加速数据分析技术的商业应用。2014年年初，新加坡资讯通信发展管理局(IDA)还聘请了首任首席数据科学家，专门推进政府数据的开放和价值开发。

从以上信息显示，数据分析不仅在各国政府及在各种各样的公司里占据了主要地位，而且伴随着计算科学的发展，从小型创业公司到专业的数据分析公司，数据分析行业都获得了巨大的发展。可以说，数据分析技术是一把让企业通向成功之门的金钥匙。

三、大数据国内发展

2012年8月国务院制定了促进信息消费扩大内需的文件，推动商业企业加快信息基础设施演进升级，增强信息产品供给能力，形成行业联盟，制定行业标准，构建大数据产业链，促进创新链与产业链有效嫁接。2015年9月，国务院公开发布《国务院关于印发促进大数据发展行动纲要的通知》(简称为《促进大数据发展行动纲要》)。《促进大数据发展行动纲要》提出“2017年年底前形成跨部门数据资源共享共用格局”。同时，开启大众创业、万众创新的创新驱动新格局。形成公共数据资源合理适度开放共享的法规制度和政策体系，2018年年底前建成国家政府数据统一开放平台，率先在信用、交通、医疗、卫生、就业、社保、地理、文化、教育、科技、资源、农业、环境、安监、金融、质量、统计、气象、海洋、企业登记监管等重要领域实现公共数据资源合理适度向社会开放，带动社会公众开展大数据增值性、公益性开发和创新应用，充分释放数据红利，激发大众创业、万众创新的活力”。

同时，构建大数据研究平台，整合创新资源，实施“专项计划”，突破关键技术。大力推进国家发改委和中科院基础研究大数据服务平台应用示范项目，广东率先启动大数据战略推动政府转型，北京正积极探索政府公布大数据供社会开发，上海也启动大数据研发三年行动计划。当前，在政府部门数据对外开放，由企业系统分析大数据进行投资经营方面，上海无疑是先行一步。2014 年 5 月 15 日，上海市开始推动各级政府部门将数据对外开放，并鼓励社会对其进行加工和运用。根据上海市经济和信息化委员会(简称“经信委”)印发的《2014 年度上海市政府数据资源向社会开放工作计划》，目前已确定 190 项数据内容作为 2014 年重点开放领域，涵盖 28 个市级部门，涉及公共安全、公共服务、交通服务、教育科技、产业发展、金融服务、能源环境、健康卫生、文化娱乐等领域。其中市场监管类数据和交通数据资源的开放将成为重点，这些与市民息息相关的信息查询届时将完全开放。这意味着企业运用大数据在上海“掘金”时代的来临，企业投资和上海民生相关的产业，如交通运输、餐饮等，可以不再“盲人摸象”。在立足国家战略和产业政策推动大数据收集和分析技术快速发展的同时，我们也应清醒地认识到避免数据垄断和保护数据安全的重要性，及早开展相关法律法规的探讨和研究。伴随着大数据时代的来临，世界各国对数据的重视提到了前所未有的高度。套上大数据的光环后，原本那些存放在服务器里平淡无奇的陈年旧数一夜之间身价倍增。按照世界经济论坛报告的看法，“大数据为新财富，价值堪比石油”。正如大数据之父维克托·迈尔·舍恩伯格所预测，“虽然数据还没有被列入企业的资产负债表，但这只是一个时间问题”。

今天的国家将大数据视为国家战略，并且在实施上，也已经进入企业战略层面，这种认识已经远远超出当年的信息化战略。但是，大数据是信息化时代的“石油”，开发大数据资源的能力将影响未来国家的核心竞争力。中国不能幻想走别人修好的道路，更不能靠等，只能依赖自身的能力加速前行，这种能力就是将数据转化为信息和知识的速度与技术，而这种转化速度和技术，则决定了大数据技术能力的高低。

从 2003 年年底信息产业部电子行业职业技能鉴定指导中心(现为“工业和信息化部教育与考试中心”)正式设立“数据分析师”培训项目，并制定出数据分析师培训、考试及管理办法。到 2014 年，中国的数据分析行业已经走过了 11 个年头。这期间中国的数据分析师、数据分析师事务所、行业协会从无到有，发展越来越快，业务领域也从最初的投资数据分析逐步转向经营数据分析。

2014 年 11—12 月期间，中国商业联合会数据分析专业委员会(China Data Analysis Committee，China General Chamber of Commerce，缩写 CDAC，以下简称“协会”)对全国数据分析行业进行了一次全国范围的调研。根据调研情况，协会发布了《2014 年度中国数据分析行业年度发展报告》(以下简称“报告”)，这是协会连续五年对国内数据分析行业进行调研并发布的行业白皮书。2014 年是中国数据分析行业第二个十年发展的开局之年，随着整个社会对数据价值的认识越来越深刻，随着数据产业链上不同领域的技术进步和资本投入，中国的数据分析行业有着此阶段的发展特点。

目前国家将大数据视为国家战略，并且在实施上，也已经进入企业战略层面。大数据是信息化时代的“石油”，开发大数据资源和数据分析能力将影响未来国家的核心竞争力——这种能力就是将数据转化为信息和知识的速度与技术，而这种转化速度和技术，则决定了大数据技术能力的高低。但这里一定要避免“大数据”=“技术开发”的误区，技术发展到什么程度，只有一小部分是由科学家追求极致的精神驱动，大部分原因是因为业务发展到一定程度，要求技术必须做出进步才能达成目标的。

四、大数据未来的发展趋势

数据生态系统逐渐丰富并影响企业商业模式。数据分析行业的整个产业链可以从基础数据的采集开始定义，包括数据采集、数据存储、数据处理(含数据清洁)、数据分析，直到数据分析结果的呈现和商业应用。通过对市场的了解和判断，我们认为目前国内整个数据分析产业链的布局相对完整，但局部环节的竞争程度差异化明显。

在数据采集领域，综合型大数据源市场处于结构化整合阶段，数据源平台企业进行以数据资产为核心的整合，最终会形成几大平台型的数据源企业，或是纯粹的用户场景数据平台。而垂直型大数据源市场处于布局阶段，全新行业细分市场的用户场景数据企业诞生，目前还处于积累用户量的阶段，不断有新的企业成长起来，抢夺已有的入口或开辟全新的入口；数据存储及挖掘市场结构较为稳定，国际巨头垄断，寡头格局已经形成，国内企业短期内很难超越。据相关数据显示，在硬盘存储器领域，EMC、IBM、NetApp 等前五家厂商在全球的市场占有率合计超过 70%；在服务器领域，HP、IBM、Dell 等前五家厂商市场占有率超过 85%；在数据库软件领域集中度更高，Oracle、IBM、Microsoft 等前五家厂商超过 90%。在数据应用市场，国内企业机会较多，但技术仍不成熟。其中行业应用主要包括商业智能、企业大数据应用系统和在线数据分析平台等，辅助应用包括数据可视化、视频、语音识别等通用性的大数据应用。

在全球经济、技术一体化的今天，我国 IT 行业已经开启了大数据的起航之旅，大数据已经在经济领域发挥重要作用。截至 2014 年，政府、互联网、电信、金融等领域的市场规模占据近一半的市场份额。大数据在主要经济领域的发展趋势如下介绍。

1. 大数据在经济预警方面发挥重要作用

在 2008 年金融危机中，阿里巴巴平台的海量交易记录预测了经济指数的下滑。2008 年年初，阿里巴巴平台上整个买家询盘数急剧下滑，预示了经济危机的来临。数以万计的中小制造商及时获得阿里巴巴的预警，为预防危机做好了准备。

2. 大数据分析成为市场营销的重要手段

与传统的市场研究方法不同，大数据的市场研究方法不再局限于抽样调查，而是基于几乎全样本空间。例如，百度拥有中国最大的消费者行为数据库，覆盖 95% 的中国网民，搜索市场占比达 87%。百度基于最真实的用户行为数据和多维度研究工具，帮助宝洁精准

地定位了消费者的地域分布、兴趣爱好等信息，根据百度分析的结论，宝洁适时地调整了营销策略。

3. 大数据在临床诊断、远程监控、药品研发等领域发挥重要作用

目前我国已经有十余座城市开展了数字医疗。病历、影像、远程医疗等都会产生大量的数据并形成电子病历及健康档案。基于这些海量数据，医院能够精准地分析病人的体征、治疗费用和疗效数据，可避免过度及副作用较为明显的治疗，此外还可以利用这些数据进行计算机远程监护，对慢性病进行管理等。

4. 大数据为金融领域的客户管理、营销管理及风险管理提供重要支撑

大数据能够解决金融领域海量数据的存储、查询优化，以及声音、影像等非结构化数据的处理。金融系统可以通过大数据分析平台，导入客户社交网络、电子商务、终端媒体产生的数据，从而构建客户视图。依托大数据平台可以进行客户行为跟踪、分析，进而获取用户的消费习惯、风险收益偏好等。针对用户这些特性，银行等金融部门能够实施风险及营销管理。

当前，我国正处在全面建设小康社会的征程中，工业化、信息化、城镇化、农业现代化任务很重，建设下一代信息基础设施，发展现代信息技术产业体系，健全信息安全保障体系，推进信息网络技术广泛运用，是实现四化同步发展的保证。大数据分析对我们深刻领会世情和国情，把握规律，实现科学发展，做出科学决策具有重要意义。

中国人口居世界首位，将会成为产生数据量最多的国家，但我们对数据保存不够重视，对存储数据的利用率也不高。此外，我国一些部门和机构拥有大量数据却不愿与其他部门共享，导致信息不完整或重复投资。政府应通过体制、机制改革打破数据割据与封锁，应注重公开信息，重视数据挖掘。美国联邦政府已建立统一数据开放门户网站，为社会提供信息服务并鼓励挖掘与利用。

1.1.2 数据分析隐藏的风险和困境

一、用户隐私

互联网时代扑面而来，网络提速，也渐渐地让每个人都享受到了互联网带来的好处，不管是手机、笔记本、台式机、平板电脑，还是电子书，这些设备都让我们与互联网紧密相连，平均个人在线时间已经远远超出了过去几年的平均水平。上网带来方便快捷的同时也带来了相关问题。人们通过互联网观看视频、网上购物、参加活动，都要涉及用户注册、用户密码、通信地址、电话、电子邮件地址等。这些个人信息很有可能被用来进行非法活动。如何保护这些个人信息不被泄漏和滥用是一个大家都在关注的问题。数据分析也同样面临着这样的困境。印度在2005年颁布了《个人数据与数据分析保护法》。希望用法律来保护个人隐私，明确了数据分析和个人隐私之间的法律权限。