

大数据下图书馆信息服务模式研究

王军光 马芳 冯宏宇 著

吉林文史出版社

大数据下图书馆信息服务模式研究

王军光 马芳 冯宏宇 著

吉林文史出版社

图书在版编目 (CIP) 数据

大数据下图书馆信息服务模式研究 / 王军光, 马芳,
冯宏宇著. -- 长春: 吉林文史出版社, 2016.10
ISBN 978-7-5472-3603-1

I. ①大… II. ①王… ②马… ③冯… III. ④图书馆
工作—情报服务—研究 IV. ① G251

中国版本图书馆 CIP 数据核字 (2016) 第 261644 号

大数据下图书馆信息服务模式研究

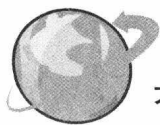
著 者	王军光 马芳 冯宏宇
出 版 人	孙建军
责任编辑	陈春燕 张蕊
封面设计	曹超
出版发行	吉林文史出版社有限责任公司
地 址	长春市人民大街 4646 号
网 址	www.jlws.com.cn
制 作	山东天图文化传媒有限公司
印 刷	南昌市彩艺印刷有限公司
开 本	787x1092mm 1/16
印 张	14
字 数	240 千字
版 次	2017 年 7 月第 1 版 2017 年 7 月第一次印刷
书 号	ISBN 978-7-5472-3603-1
定 价	48.00 元

前 言

随着移动互联网技术的不断发展，大数据给高校图书馆带来前所未有的机遇和挑战，如今受互联网、云计算、物联网、社交网络等技术发展数据的产生呈现出爆炸式增长，一系列信息数据的出现意味着大数据时代的来临。为了能更准确、更及时的把信息传递给用户，图书馆工作人员应该及时调整服务模式，采用先进的大数据技术来对图书馆信息进行适时的处理。把结构复杂、形式多变的大数据变成有价值的信息，然后供用户使用。将大数据和图书馆信息服务联系起来成为了图书馆发展的趋势，将大数据技术融入图书馆信息服务，为用户提供个性化服务也成为了图书馆发展的必然趋势。《大数据下图书馆信息服务模式研究》正是基于当前这种大数据的背景，针对国内图书馆服务存在的共性问题和不足展开分析，通过对多所高校图书馆创新信息服务不同案例进行分析，探讨了大数据对高校图书馆信息服务的影响，研究了基于大数据的高校图书馆信息服务内容以及大数据环境下构建高校图书馆信息服务创新模式。并探讨了大数据思想将会给图书馆服务带来有效改善，具有较强的现实意义和应用价值。所以我们必须紧紧抓住图书馆大发展，在图书馆逐步开展大数据技术的研究和应用工作，顺应时代发展要求，不断满足读者的需求，实现图书馆顺应时代与时俱进的飞速发展。

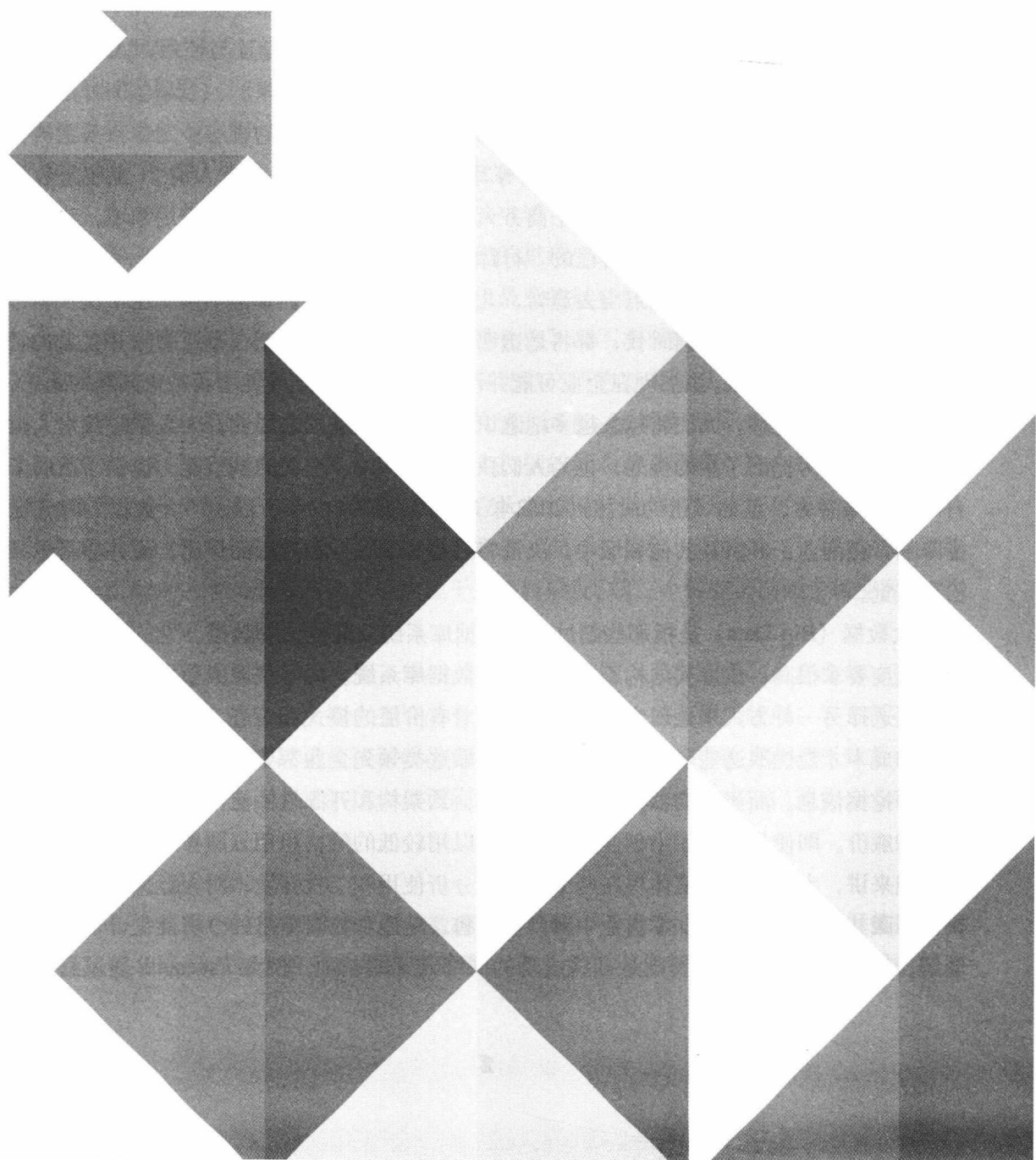
目 录

第一章 大数据.....	1
第一节 大数据概念.....	6
第二节 大数据的处理技术.....	11
第三节 大数据特点.....	15
第四节 大数据核心价值.....	25
第二章 大数据有助于挖掘市场机会.....	33
第一节 大数据与企业决策能力.....	38
第二节 大数据与变革商业模式.....	44
第三节 大数据有助于驱动智慧和谐社会.....	51
第四节 大数据有助于预判未来趋势.....	58
第三章 国内外大数据研究发展现状.....	63
第一节 大数据渗透于全球社会经济文化各领域.....	69
第二节 欧美等发达地区大数据国家战略.....	76
第三节 我国大数据国家战略.....	81
第四章 高校图书馆与传统信息服务.....	85
第一节 高校图书馆概述.....	90
第二节 高校图书馆传统信息服务.....	95
第三节 高校图书馆读者服务宗旨.....	99
第四节 大数据带来高校图书馆思维变革.....	104
第五节 大数据对高校图书馆的影响.....	109



第六节 国外大数据推进图书馆应用的实践	115
第五章 大数据时代高校图书馆信息服务创新内容	119
第一节 高校图书馆大数据整合系统平台	122
第二节 高校图书馆大数据环境特点与价值可用性定位	124
第三节 高校图书馆大数据可用性面临的挑战与保障	129
第四节 高校图书馆大数据资源共享	134
第五节 高校图书馆大数据检索服务	138
第六章 大数据时代高校图书馆知识服务	143
第一节 高校图书馆咨询服务新模式	147
第二节 高校图书馆知识咨询服务驱动因素	153
第三节 大数据时代高校图书馆阅读推广	161
第七章 大数据时代图书馆信息创新服务案例	165
第一节 线下到线上存取	169
第二节 “云计算”在图书馆的应用	175
第三节 智慧图书馆	182
第四节 大数据在图书馆自助服务与移动服务中的应用	186
第五节 RFID 与手机二维码在图书馆的应用	193
第六节 大数据时代图书馆信息素养教育创新及宣传推广	201
第八章 大数据时代高校图书馆信息服务展望	207
第一节 大数据时代高校图书馆信息服务面临的问题	210
第二节 大数据时代提升高校图书馆信息服务应对策略	212
结束语	217
参考文献	218

第一章 大数据





人类世界,有很多个“时代”。如原始社会时代、奴隶主时代、封建帝国时代、资本主义、社会主义时代。信息时代是我们目前所处的时代。在这个时代,信息(也是数据)极大膨胀和爆炸,因此诞生了“大数据时代”。在这个时代,数据的处理、加工、生产、流通、管理成为了数据人必不可少的一部分。是生活,也是工作,更是娱乐。数据是人的一部分,人也是数据的一部分。可以说,人类在这个“大数据时代”,任何行为、任何事物、任何人类信息都被数据化、电子化了。云计算、云存储是应对数据大膨胀而提出的数据存储、管理、计算所提出的优化解决方案。而物联网则是将人类行为、物品行为信息收集起来,存放在网络中的一种终端解决方案。不管是哪一个解决方案,都是将人类世界信息化、数据化、电子化进行到底的解决方案。

进入 2012 年以来,大数据(Big Data)一词越来越多地被提及与使用,人们用它来描述和定义信息爆炸时代产生的海量数,它已经出现在《纽约时报》、《华尔街时报》的专栏封面,进入美国白宫网的新闻,现身在国内一些互联网主题的讲座沙龙中,甚至被嗅觉灵敏的国君证券、国泰君安、银河证券等写进了投资推荐报告。有人说 21 世纪是数据信息时代,移动互联、社交网络、电子商务大大拓展了互联网的疆界和应用领域。我们在享受便利的同时,也无偿贡献了自己的“行踪”。现在互联网不但知道对面是一只狗,还知道这只狗喜欢什么食物,几点出去遛弯,几点回窝睡觉。我们不得不接受这个现实,每个人在互联网进入大数据时代,都将是透明性存在。各种数据正在迅速膨胀并变大,它决定着企业的未来发展,虽然现在企业可能并没有意识到数据爆炸性增长带来问题的隐患,但是随着时间的推移,人们将越来越多地意识到数据对企业的重要性。大数据时代对人类的数据驾驭能力提出了新的挑战,也为人们获得更为深刻、全面的洞察能力提供了前所未有的空间与潜力。正如《纽约时报》2012 年 2 月的一篇专栏中所称,“大数据”时代已经降临,在商业、经济及其他领域中,决策将日益基于数据和分析而作出,而并非基于经验和直觉。

大数据(Big Data)是指那些超过传统数据库系统处理能力的数据。它的数据规模和转输速度要求很高,或者其结构不适合原本的数据库系统。为了获取大数据中的价值,我们必须选择另一种方式来处理它。数据中隐藏着有价值的模式和信息,在以往需要相当的时间和成本才能提取这些信息。如沃尔玛或谷歌这类领先企业都要付高昂的代价才能从大数据中挖掘信息。而当今的各种资源,如硬件、云架构和开源软件使得大数据的处理更为方便和廉价。即使是在车库中创业的公司也可以用较低的价格租用云服务时间了。对于企业组织来讲,大数据的价值体现在两个方面:分析使用和二次开发。对大数据进行分析能揭示隐藏其中的信息。例如零售业中对门店销售、地理和社会信息的分析能提升对客户理解。对大数据的二次开发则是那些成功的网络公司的长项。例如 Facebook 通过结合大

量用户信息，定制出高度个性化的用户体验，并创造出一种新的广告模式。这种通过大数据创造出新产品和服务的商业行为并非巧合，谷歌、雅虎、亚马逊和 Facebook 它们都是大数据时代的创新者。

数据大时代的特征：

一、数据极多。大到需要云存储、云计算解决方案来加以优化存储管理、数据计算处理。

二、数据包含领域广。包含从生产、消费、工作、学习、生活、政治所有领域的的数据，这些数据产生非常迅速、不断推陈出新。

三、数据影响范围大。数据不仅仅是供人查阅分享，更多的是进一步影响了人们的生活、工作、学习娱乐环境等。

四、数据化、信息化将席卷整个人类世界。最终的效果是：人类世界自动化、电子化、数据信——信息化。

大数据时代对生活、工作的影响

维克托·尔耶·舍恩伯格在《大数据时代：生活、工作与思维的大变革》中前瞻性地指出，大数据带来的信息风暴正在变革我们的生活、工作和思维，大数据开启了一次重大的时代转型，并用三个部分讲述了大数据时代的思维变革、商业变革和管理变革。维克托最具洞见之处在于，他明确指出，大数据时代最大的转变就是放弃对因果关系的渴求而取而代之关注相关关系。也就是说只要知道“是什么”而不需要知道“为什么”。这颠覆了千百年来人类的思维惯例，对人类的认知和与世界交流的方式提出了全新的挑战。

大数据是人们获得新的认知，创造新的价值的源泉；大数据还是改变市场、组织机构，以及政府与公民关系的方法。维克托认为，大数据的核心就是预测。这个核心代表着我们分析信息时的三个转变。第一个转变就是，在大数据时代我们可以分析更多的数据，有时候甚至可以处理和某个特别现象相关的所有数据，而不再依赖于随机采样。第二个改变就是研究数据如此之多以至于我们不再热衷于追求精确度。第三个转变因前两个转变而促成，即我们不再热衷于寻找因果关系。

近些年，大数据已经和云计算一样，成为时代的话题。大数据是怎么产生的，商业机会在哪？研究机会在哪？这个概念孕育着一个怎样的未来？企业如何应对？一个好的企业应该未雨绸缪，从现在开始就应该着手准备，为企业的后期的数据收集和分析做好准备，企业可以从下面五个方面着手，这样当面临铺天盖地的大数据的时候，以确保企业能够快速发展，具体为下面五点：

（一）以企业的数据为目标

几乎每个组织都可能源源不断的数据需要收集，无论是社交网络还是车间传感器设备，而且每个组织都有大量的数据需要处理，IT 人员需要了解自己企业运营过程中都产



生了什么数据，以自己的数据为基准，确定数据的范围。

（二）以业务需求为准则

虽然每个企业都会产生大量数据，而且互不相同、多种多样的，这就需要企业 IT 人员现在开始收集确认什么数据是企业业务需要的，找到最能反映企业业务情况的数据。

（三）重新评估企业基础设施

大数据需要在服务器和存储设施中进行收集，并且大多数的企业信息管理体系结构将会发生重要大变化，IT 经理则需要准备扩大他们的系统，以解决数据的不断扩大，IT 经理要了解公司现有 IT 设施的情况，以组建处理大数据的设施为导向，避免一些不必要的设备的购买。

（四）重视大数据技术

大数据是最近几年才兴起的词语，而并不是所有的 IT 人员对大数据都非常了解，例如 Hadoop, MapReduce, NoSQL 等技术都是近年刚兴起的技术，企业 IT 人员要多关注这方面的技术和工具，以确保将来能够面对大数据的时候做出正确的决定。

（五）培训企业的员工

当大数据到临的时候，企业将会缺少这方面的采集收集分析方面的人才，对于一些公司，工作人员面临大数据将是一种挑战，企业要在平时多对员工进行这方面的培训，以确保在大数据到来时，员工也能适应相关的工作。

做到上面的几点，当大数据时代来临的时候，面临大量数据将不是束手无策，而是成竹在胸，而从数据中得到的好处也将促进企业快速发展。

大数据时代的发展方向、趋势

第一大趋势：应用软件泛互联网化。所谓泛互联网化，就是指应用软件都会和互联网联通，成为用户接入互联网，享用网络服务的媒介。

第二大趋势：行业应用的垂直整合。在这个趋势下，越靠近终端用户公司，在产业链中拥有越大的发言权。

第三大趋势数据将成为资产。未来企业的竞争，将是拥有数据规模和活性的竞争，将是对数据解释和运用的竞争。

大数据的时代已然来临。IDC 数据显示，在 2006 年全世界的电子数据存储量为 18 万 PB，而如今这个数字已经达到 180 万 PB，短短 5 年间就已经增长了一个数量级。而根据预测，2015 年这个数字则会达到如同天文数字般的 800 万 PB。就在此时此刻，海量数据依然源源不断地产生，从不停息。面对这些“大数据”有些人叹息抱怨，害怕数据量的剧增对于现有 IT 架构的冲击；有些人积极主动探寻应对海量数据的应对与解决之道；还有一些人则是顺势而为抓住时代发展的商业机会成为富有活力的创新者。大数据的发展催生了诸多

商业机会和商业模式。而这些公司所面对的独特时代背景，就注定了它们必会受到市场和资本的追捧。它们中的一些或是已经融资成功，进入高速发展期；或是被成功收购，帮助投资人和创始人成功从项目中退出。而很多上市公司，也开始在这一领域动作频繁，积极布局，这也从侧面反应了这一领域的广阔前景和巨大的利润空间。



第一节 大数据概念

大数据 (big data), 指无法在可承受的时间范围内用常规软件工具进行捕捉、管理和处理的数据集合, 是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力来适应海量、高增长率和多样化的信息资产在维克托·迈尔·舍恩伯格及肯尼斯·库克耶编写的《大数据时代》中大数据指不用随机分析法 (抽样调查) 这样的捷径, 而采用所有数据进行分析处理。大数据的 5V 特点 (IBM 提出): Volume (大量)、Velocity (高速)、Variety (多样)、Value (价值)、Veracity (真实性)。

应用学科: 计算机、信息科学、统计学

适用领域范围: BI、工业 4.0、云计算、物联网、互联网+、人工智能

大数据技术的战略意义不在于掌握庞大的数据信息, 而在于对这些含有意义的数据进行专业化处理。换言之, 如果把大数据比作一种产业, 那么这种产业实现盈利的关键, 在于提高对数据的“加工能力”, 通过“加工”实现数据的“增值”。从技术上看, 大数据与云计算的关系就像一枚硬币的正反面一样密不可分。大数据必然无法用单台的计算机进行处理, 必须采用分布式架构。它的特色在于对海量数据进行分布式数据挖掘。但它必须依托云计算的分布式处理、分布式数据库和云存储、虚拟化技术。随着云时代的来临, 大数据 (Big data) 也吸引了越来越多的关注。《著云台》的分析师团队认为大数据 (Big data) 通常用来形容一个公司创造的大量非结构化数据和半结构化数据, 这些数据在下载至关系型数据库用于分析时会花费过多时间和金钱。大数据分析常和云计算联系到一起, 因为实时的大型数据集分析需要像 MapReduce 一样的框架来向数十、数百甚至数千的电脑分配工作。大数据需要特殊的技术, 以有效地处理大量的容忍经过时间内的数据。适用于大数据的技术, 包括大规模并行处理 (MPP) 数据库、数据挖掘电网、分布式文件系统、分布式数据库、云计算平台、互联网和可扩展的存储系统。

现在的社会是一个高速发展的社会, 科技发达, 信息流通, 人们之间的交流越来越密切, 生活也越来越方便, 大数据就是这个高科技时代的产物。阿里巴巴创办人马云来台演讲中就提到, 未来的时代将不是 IT 时代, 而是 DT 的时代, DT 就是 Data Technology 数据科技, 显示大数据对于阿里巴巴集团来说举足轻重。有人把数据比喻为蕴藏能量的煤矿。煤炭按照性质有焦煤、无烟煤、肥煤、贫煤等分类, 而露天煤矿、深山煤矿的挖掘成本又不一样。与此类似大数据并不在“大”, 而在于“有用”。价值含量、挖掘成本比数量更为重要。对于很多行业而言, 如何利用这些大规模数据是成为赢得竞争的关键。

大数据的价值体现在以下几个方面:

一、对大量消费者提供产品或服务的企业可以利用大数据进行精准营销；

二、做小而美模式的中长尾企业可以利用大数据做服务转型；

三、面临互联网压力之下必须转型的传统企业需要与时俱进充分利用大数据的价值。

不过，“大数据”在经济发展中的巨大意义并不代表其能取代一切对于社会问题的理性思考，科学发展的逻辑不能被湮没在海量数据中。著名经济学家路德维希·冯·米塞斯曾提醒过：“就今日言，有很多人忙碌于资料之无益累积，以致对问题之说明与解决，丧失了其对特殊的经济意义的了解。”这确实是需要警惕的。在这个快速发展的智能硬件时代，困扰应用开发者的一个重要问题就是如何在功率、覆盖范围、传输速率和成本之间找到那个微妙的平衡点。企业组织利用相关数据和分析可以帮助它们降低成本、提高效率、开发新产品、做出更明智的业务决策等等。例如，通过结合大数据和高性能的分析，下面这些对企业有益的情况都可能会发生：

一、及时解析故障、问题和缺陷的根源，每年可能为企业节省数十亿美元。

二、为成千上万的快递车辆规划实时交通路线，躲避拥堵。

三、分析所有 SKU，以利润最大化为目标来定价和清理库存。

四、根据客户的购买习惯，为其推送他可能感兴趣的优惠信息。

五、从大量客户中快速识别出金牌客户。

六、使用点击流分析和数据挖掘来规避欺诈行为。

大数据就是互联网发展到现今阶段的一种表象或特征而已，没有必要神话它或对它保持敬畏之心，在以云计算为代表的技术创新大幕的衬托下，这些原本很难收集和使用的数据开始容易被利用起来了，通过各行各业的不断创新，大数据会逐步为人类创造更多的价值。其次，想要系统地认知大数据，必须要全面而细致地分解它，我着手从三个层面来展开：

第一层面是理论，理论是认知的必经途径，也是被广泛认同和传播的基线。在这里从大数据的特征定义理解行业对大数据的整体描绘和定性；从对大数据价值的探讨来深入解析大数据的珍贵所在；洞悉大数据的发展趋势；从大数据隐私这个特别而重要的视角审视人和数据之间的长久博弈。

第二层面是技术，技术是大数据价值体现的手段和前进的基石。在这里分别从云计算、分布式处理技术、存储技术和感知技术的发展来说明大数据从采集、处理、存储到形成结果的整个过程。

第三层面是实践，实践是大数据的最终价值体现。在这里分别从互联网的大数据，政府的大数据，企业的大数据和个人的大数据四个方面来描绘大数据已经展现的美好景象及即将实现的蓝图。洛杉矶警察局和加利福尼亚大学合作利用大数据预测犯罪的发生。google 流感趋势 (Google Flu Trends) 利用搜索关键词预测禽流感的散布。统计学家内特·西



尔弗 (Nate Silver) 利用大数据预测 2012 美国选举结果。麻省理工学院利用手机定位数据和交通数据建立城市规划。梅西百货的实时定价机制。根据需求和库存的情况, 该公司基于 SAS 的系统对多达 7300 万种货品进行实时调价。医疗行业早就遇到了海量数据和非结构化数据的挑战, 而近年来很多国家都在积极推进医疗信息化发展, 这使得很多医疗机构有资金来做大数据分析。

趋势:

趋势一: 数据的资源化

何为资源化, 是指大数据成为企业和社会关注的重要战略资源, 并已成为大家争相抢夺的新焦点。因而, 企业必须要提前制定大数据营销战略计划, 抢占市场先机。

趋势二: 与云计算的深度结合

大数据离不开云处理, 云处理为大数据提供了弹性可拓展的基础设备, 是产生大数据的平台之一。自 2013 年开始, 大数据技术已开始和云计算技术紧密结合, 预计未来两者关系将更为密切。除此之外, 物联网、移动互联网等新兴计算形态, 也将一齐助力大数据革命, 让大数据营销发挥出更大的影响力。

趋势三: 科学理论的突破

随着大数据的快速发展, 就像计算机和互联网一样, 大数据很有可能是新一轮的技术革命。随之兴起的数据挖掘、机器学习和人工智能等相关技术, 可能会改变数据世界里的很多算法和基础理论, 实现科学技术上的突破。

趋势四: 数据科学和数据联盟的成立

未来, 数据科学将成为一门专门的学科, 被越来越多的人所认知。各大高校将设立专门的数据科学类专业, 也会催生一批与之相关的新的就业岗位。与此同时, 基于数据这个基础平台, 也将建立起跨领域的数据共享平台, 之后数据共享将扩展到企业层面, 并且成为未来产业的核心一环。

趋势五: 数据泄露泛滥

未来几年数据泄露事件的增长率也许会达到 100%, 除非数据在其源头就能够得到安全保障。可以说在未来每个财富 500 强企业都会面临数据攻击, 无论他们是否已经做好安全防范。而所有企业无论规模大小都需要重新审视今天的安全定义。在财富 500 强企业中, 超过 50% 将会设置首席信息安全官这一职位。企业需要从新的角度来确保自身以及客户数据, 所有数据在创建之初便需要获得安全保障, 而并非在数据保存的最后一个环节, 仅仅加强后者的安全措施已被证明于事无补。

趋势六: 数据管理成为核心竞争力

数据管理成为核心竞争力, 直接影响财务表现。当“数据资产是企业核心资产”的概

念深入人心之后，企业对于数据管理便有了更清晰的界定，将数据管理作为企业核心竞争力，持续发展，战略性规划与运用数据资产，成为企业数据管理的核心。数据资产管理效率与主营业务收入增长率、销售收入增长率显著正相关；此外，对于具有互联网思维的企业而言，数据资产竞争力所占比重为 36.8%，数据资产的管理效果将直接影响企业的财务表现。

趋势七：数据质量是 BI（商业智能）成功的关键

采用自助式商业智能工具进行大数据处理的企业将会脱颖而出。其中要面临的一个挑战是，很多数据源会带来大量低质量数据。想要成功，企业需要理解原始数据与数据分析之间的差距，从而消除低质量数据并通过 BI 获得更佳决策。

趋势八：数据生态系统复合化程度加强

大数据的世界不只是一个单一的、巨大的计算机网络，而是一个由大量活动构件与多元参与者元素所构成的生态系统，终端设备提供商、基础设施提供商、网络服务提供商、网络接入服务提供商、数据服务使能者、数据服务提供商、触点服务、数据服务零售商等等一系列的参与者共同构建的生态系统。而今，这样一套数据生态系统的基本雏形已然形成，接下来的发展将趋向于系统内部角色的细分，也就是市场的细分；系统机制的调整，也就是商业模式的创新；系统结构的调整，也就是竞争环境的调整等等，从而使得数据生态系统复合化程度逐渐增强。

国务院印发行动纲要促进大数据加快发展

经李克强总理签批，2015 年 9 月，国务院印发《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。《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，培育新兴业态，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研创新有机结合，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。

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。2015 年 9 月 18 日贵州省启动我国首个大数据综合试验区的建设工作，力争通



过3至5年的努力,将贵州大数据综合试验区建设成为全国数据汇聚应用新高地、综合治理示范区、产业发展聚集区、创业创新首选地、政策创新先行区。围绕这一目标,贵州省将重点构建“三大体系”,重点打造“七大平台”,实施“十大工程”。

“三大体系”是指构建先行先试的政策法规体系、跨界融合的产业生态体系、防控一体的安全保障体系;“七大平台”则是指打造大数据示范平台、大数据集聚平台、大数据应用平台、大数据交易平台、大数据金融服务平台、大数据交流合作平台和大数据创业创新平台;“十大工程”即实施数据资源汇聚工程、政府数据共享开放工程、综合治理示范提升工程、大数据便民惠民工程、大数据三大业态培育工程、传统产业改造升级工程、信息基础设施提升工程、人才培养引进工程、大数据安全保障工程和大数据区域试点统筹发展工程。此外,贵州省将计划通过综合试验区建设,探索大数据应用的创新模式,培育大数据交易新的做法,开展数据交易的市场试点,鼓励产业链上下游之间的数据交换,规范数据资源的交易行为,促进形成新的业态。国家发展改革委有关专家表示,大数据综合试验区建设不是简单的建产业园、建数据中心、建云平台等,而是要充分依托已有的设施资源,把现有的利用好,把新建的规划好,避免造成空间资源的浪费和损失。探索大数据应用新的模式,围绕有数据、用数据、管数据,开展先行先试,更好地服务国家大数据发展战略。

大数据的 IT 分析

大数据概念应用到 IT 操作工具产生的数据中,大数据可以使 IT 管理软件供应商解决大广泛的业务决策。IT 系统、应用和技术基础设施每天每秒都在产生数据。大数据非结构化或者结构数据都代表了“所有用户的行为、服务级别、安全、风险、欺诈行为等更多操作”的绝对记录。大数据分析的产生旨在于 IT 管理,企业可以将实时数据流分析和历史相关数据相结合,然后大数据分析并发现它们所需的模型。反过来,帮助预测和预防未来运行中断和性能问题。进一步来讲,他们可以利用大数据了解使用模型以及地理趋势,进而加深大数据对重要用户的洞察力。他们也可以追踪和记录网络行为,大数据轻松地识别业务影响;随着对服务利用的深刻理解加快利润增长;同时跨多系统收集数据发展 IT 服务目录。大数据分析的想法,尤其在 IT 操作方面,大数据对于我们发明并没有什么作用,但是我们一直在其中。Gartner 已经关注这个话题很多年了,基本上他们已经强调,如果 IT 正在引进新鲜灵感,他们将会扔掉大数据老式方法开发一个新的 IT 操作分析平台。

第二节 大数据的处理技术

数据是指通过 RFID 射频数据、传感器数据、社交网络交互数据及移动互联网数据等方式获得的各种类型的结构化、半结构化（或称之为弱结构化）及非结构化的海量数据，是大数据知识服务模型的根本。重点要突破分布式高速高可靠数据爬取或采集、高速数据全映像等大数据收集技术；突破高速数据解析、转换与装载等大数据整合技术；设计质量评估模型，开发数据质量技术。大数据采集一般分为大数据智能感知层：主要包括数据传感体系、网络通信体系、传感适配体系、智能识别体系及软硬件资源接入系统，实现对结构化、半结构化、非结构化的海量数据的智能化识别、定位、跟踪、接入、传输、信号转换、监控、初步处理和管理等。必须着重攻克针对大数据源的智能识别、感知、适配、传输、接入等技术。

一、大数据预处理技术

主要完成对已接收数据的辨析、抽取、清洗等操作。

（一）抽取：因获取的数据可能具有多种结构和类型，数据抽取过程可以帮助我们将这些复杂的数据转化为单一的或者便于处理的构型，以达到快速分析处理的目的。

（二）清洗：对于大数据，并不全是有价值的，有些数据并不是我们所关心的内容，而另一些数据则是完全错误的干扰项，因此要对数据通过过滤“去噪”从而提取出有效数据。

二、大数据存储及管理技术

大数据存储与管理要用存储器把采集到的数据存储起来，建立相应的数据库，并进行管理和调用。重点解决复杂结构化，半结构化和非结构化大数据管理与处理技术。主要解决大数据的可存储、可表示、可处理、可靠性及有效传输等几个关键问题。开发可靠的分布式文件系统（DFS）、能效优化的存储、计算融入存储、大数据的去冗余及高效低成本的大数据存储技术；突破分布式非关系型大数据管理与处理技术，异构数据的数据融合技术，数据组织技术，研究大数据建模技术；突破大数据索引技术；突破大数据移动、备份、复制等技术；开发大数据可视化技术。

开发新型数据库技术，数据库分为关系型数据库、非关系型数据库以及数据库缓存系统。其中，非关系型数据库主要指的是 NoSQL 数据库，分为：键值数据库、列存数据库、图存数据库以及文档数据库等类型。关系型数据库包含了传统关系数据库系统以及 NewSQL 数据库。开发大数据安全技术。改进数据销毁、透明加解密、分布式访问控制、数据审计等技术；突破隐私保护和推理控制、数据真伪识别和取证、数据持有完整性验证等技术。