

大数据时代下的 企业财务管理研究

DASHUJU SHIDAIXIA DE
QIYE CAIWU GUANLI YANJIU

姬潮心 王媛◎著



中国水利水电出版社
www.waterpub.com.cn

大数据时代下的企业 财务管理研究

姬潮心 王 媛 著



中国水利水电出版社

www.waterpub.com.cn

· 北京 ·

内 容 提 要

本书分为两部分,第一部分主要分析了大数据的基础知识,首先论述了大数据的应用架构与技术、大数据的管理和治理体系,而后分析了大数据与企业、大数据与云计算及大数据与云安全的相互关系;第二部分论述了大数据时代下企业的财务管理,首先分析了大数据对财务管理的影响,然后分别阐述了大数据时代下的企业预算管理、大数据时代下的企业投资管理、大数据时代下的企业财务决策管理、大数据时代下的企业财务战略管理及大数据时代下的企业财务预警管理。

本书可供高等院校计算机专业、财务管理专业的高年级本科生和研究生参考,也可供从事财务管理工作的人员参考阅读。

图书在版编目(CIP)数据

大数据时代下的企业财务管理研究/姬潮心,王媛
著.--北京:中国水利水电出版社,2018.6
ISBN 978-7-5170-6531-9

I. ①大… II. ①姬… ②王… III. ①企业管理—财务管理—研究 IV. ①F275

中国版本图书馆CIP数据核字(2018)第130119号

责任编辑:陈洁

封面设计:王伟

书 名	大数据时代下的企业财务管理研究 DASHUJU SHIDAI XIA DE QIYE CAIWU GUANLI YANJIU
作 者	姬潮心 王媛 著
出版发行	中国水利水电出版社 (北京市海淀区玉渊潭南路1号D座 100038) 网址: www.waterpub.com.cn E-mail: mchannel@263.net (万水) sales@waterpub.com.cn 电话: (010) 68367658 (营销中心)、82562819 (万水)
经 售	全国各地新华书店和相关出版物销售网点
排 版	北京万水电子信息有限公司
印 刷	三河市同力彩印有限公司
规 格	185mm×260mm 16开本 15印张 410千字
版 次	2018年9月第1版 2018年9月第1次印刷
印 数	0001—2000册
定 价	62.00元



凡购买我社图书,如有缺页、倒页、脱页的,本社营销中心负责调换

版权所有·侵权必究

前言

自2010年，美国数据科学家维克托·迈尔·舍恩伯格系统地提出“大数据”的概念后，从硅谷到北京，不断有人开始谈论大数据，人类也开始进入大数据时代，全球数据信息量呈现指数式爆炸增长之势，数据体量从PB级跃升至ZB级。根据国际数据公司（IDC）发布的2012年研究报告显示，预计到2020年，全球产生的数据总量将超过40ZB，将是地球所有海滩上沙粒数量的57倍。多元化的数据涌现，给人们的日常生活与工作带来了更多的机遇，同时也带来了众多挑战。如何应对这些数据给人们带来的机遇与挑战，成为人们越来越关注的话题；如何在海量的数据中进行分析挖掘，筛选有利的信息，成为人们迫切需要解决的问题；如何运用互联网信息数据开发设计软件，并使之服务于人们的生产生活，成为大数据运用的核心议题，也成为数据发展的趋势。

大数据技术的发展与应用，将对社会的组织结构、经济的运行机制、社会的生活方式、国家的治理模式、企业的决策架构、商业的业务策略以及个人的生活、工作和思维方式等产生深远的影响。企业、政府和社会逐渐渗透到互联网环境中，使人们的生活环境和生活方式向数字化、网络化方向迈进。数据逐渐被应用于各个行业领域，这成为了人类历史上又一巨大变革。

大数据对中国经济的发展也起到了巨大的推动作用，带动了技术研发、管理方式以及商业模式的创新。在大数据时代，企业可以通过分析海量的数据使得自己的决策更加正确。但大数据时代也对企业的运营与管理提出了更高的要求，如对于财务人员来说，必须要掌握更加先进的技术才能应对大数据带来的挑战。对于财务的管理者以及企业的决策者而言，更应该高瞻远瞩，应该重新审视和思考财务工作，提升自己分析、统筹数据的技能，掌握各种企业价值的评估方法，从而提高自己应对环境变化的决策能力。大数据必然会对企业的财务管理工作有很大的影响，那么在当下，企业就应该将财务决策置于大数据的环境之下进行考虑，建立基于数据的服务理念，并且及时分析当前的数据以不断修正企业决策，在大数据的背景下，企业应该对各种数据进行分析，从而作出全面的调整，继而创新自己的商业模式系统。

为了应对大数据对财务管理的新诉求，本人写作了此书。本书分为两部分，第一部分主要阐述了大数据的基础知识，包括大数据来源与发展，大数据概念、特征与类型及大数

据未来趋势，而后分析了大数据应用架构与技术、大数据管理和治理体系，最后分析了大数据与企业、大数据与云计算及大数据与云安全的相互关系；第二部分论述了大数据下企业的财务管理，首先分析了大数据对财务管理的影响，然后分别阐述了大数据时代企业预算管理、大数据时代企业投资管理、大数据时代企业财务决策管理、大数据时代企业财务战略管理及大数据时代企业财务预警管理。

本书的写作过程花费了大量时间，翻阅了大量资料，并且就有些问题咨询了相关的专家，以求提高本书的价值。但是，由于本人能力有限，本书可能还存在许多不足之处，希望广大读者批评指正。最后，诚挚地感谢在本书的写作过程中给予作者帮助的广大亲友！

作者

2018年4月

目 录

第一部分 大数据的基础知识

第 1 章 大数据概述	3
1.1 大数据的来源与发展	3
1.2 大数据的概念、特征与类型	8
1.3 大数据的未来发展趋势	13
第 2 章 大数据的应用架构与关键技术	22
2.1 大数据的应用架构	22
2.2 大数据的关键技术	28
第 3 章 大数据的管理和治理体系	40
3.1 建立数据驱动的管理体系和架构	40
3.2 大数据治理体系	42
3.3 大数据技术管理体系	49
3.4 大数据事务管理、流程管理与易用性管理	54
第 4 章 大数据与企业	59
4.1 企业面临的挑战	59
4.2 企业大数据的存储与加工	67
4.3 大数据对企业的影响	72
第 5 章 大数据与云计算	77
5.1 云计算概述	77
5.2 云计算与大数据的相关技术	81
5.3 基于云计算的大数据应用	89
第 6 章 大数据与云安全	96
6.1 大数据的信息安全策略	96

6.2 大数据时代云安全的实践与展望	105
--------------------------	-----

第二部分 基于大数据的企业财务管理研究

第7章 大数据对财务管理的影响	121
7.1 大数据对财务管理工作的影响	121
7.2 大数据与财务信息化	126
7.3 大数据时代下企业财务管理的创新	131
第8章 大数据时代下的企业预算管理	137
8.1 预算与企业战略的关系	137
8.2 预算管理的思想与思路	141
8.3 大数据对预算的影响	144
8.4 大数据时代下全面预算管理体系的构建	147
8.5 大数据时代下全面预算过程应注意的问题	151
第9章 大数据时代下的企业投资管理	154
9.1 企业投资、投资项目与项目管理	154
9.2 大数据时代下企业投资决策竞争情报的需求与服务	159
9.3 大数据在投资决策中的应用价值	163
9.4 大数据时代下企业投资决策的优化	167
第10章 大数据时代下的企业财务决策管理	173
10.1 大数据时代理论下市场的演变	173
10.2 大数据与企业竞争力	178
10.3 大数据对财务决策的影响	182
10.4 大数据时代下的企业财务决策战略目标	186
第11章 大数据时代下的企业财务战略管理	192
11.1 企业财务战略及其管理	192
11.2 大数据时代商业模式的创新	197
11.3 大数据时代企业成本的变动与控制	201
11.4 大数据时代市场的精准定位	207
第12章 大数据时代下的企业财务预警管理	211
12.1 企业财务预警效果影响因素	211
12.2 大数据时代下的企业财务预警模型设计	214
12.3 企业财务预警大数据指标数据的获取与处理	219
12.4 基于财务预警模型的企业危机管理	227
参考文献	230

第一部分 大数据的基础知识

随着经济的迅速发展，科学技术也在飞速进步，现在的人们已经步入了信息化时代。从当前社会的各方面发展进行分析，一个国家最重要的能力就是对信息的掌握能力，一个国家对于信息的掌握越清楚、越详细，那么这个国家的综合实力也不会弱。每个人对于信息的需求不同，这也表明信息的效益不同，而全部的有益信息也是经过大数据的分析得来的。海量的数据信息扩散、变动，逐步地发展成为大数据。对于“大数据”这个词语，不仅仅是形容数据的量，还用来形容处理数据的速度之快。因此，现在的大数据已经发展成为了一项广受欢迎的数据分析技术。在许多的行业都会存在着数据的影子，人们对于数据的广泛使用，标志着大数据时代的到来。大数据越来越与人类生活的方方面面联系在一起，许多的问题都利用大数据进行解决。

这一部分主要概述了大数据的基础知识，围绕大数据讲述了基本内容、大数据的应用架构与技术、大数据管理和治理体系、大数据与企业、大数据与云计算、大数据与云安全，从多个方面对大数据展开了概述。

第 1 章 大数据概述

21 世纪，大数据时代已经慢慢地到来。大数据的特点是种类繁多、规模庞大、信息生成速度快，并且具有极高的价值。本章主要是对大数据作一个概括，追溯了大数据的来源与发展，讲述了大数据的概念、特征与类型，并且在当前大数据的发展情况下，对大数据的未来趋势进行了描述。

1.1 大数据的来源与发展

一、大数据的来源

(一) 大数据的产生来源

在人类历史上，从未有哪一个时代像今天这样拥有巨大的数据，取之不尽，用之不竭，并且这些数据的产生没有限制，并不会像过去一样受到时间和空间的约束。大数据的产生，无论是对于科学技术的进步，还是对于信息技术的发展，都是值得庆祝的事情。从开始采用数据库作为数据管理的主要方式开始，人类社会的数据演变经历了三个阶段，即被动、主动和自动的阶段，由此产生了最终的大数据。

1. 运营式系统阶段

起初，数据库的产生就降低了数据管理的难度，也就是说，数据库基本上都是被运营式系统所采用，以此来当作运营式系统的数据管理子系统。例如，超市里面的销售记录系统、银行的交易记录系统、医院病人的医疗记录等。对于人类社会来说，这个阶段的数据量是在数据库的基础上产生的，因为数据库的管理，让巨大的数据得以运营。这个阶段的主要特点是，数据都会跟随着一定的运营活动而产生并记录在数据库之中，如超市的售货记录，在超市里每售出一件产品，数据库中都会记录相应的数据，不过这种数据产生的方式是被动的。

2. 用户原创内容阶段

在运营式系统阶段之后，由于互联网的诞生，让人类社会的数据量得到了第二次飞跃。在这个阶段，科学技术迅猛发展，出现了新型的社交网络软件（微博、微信等），微博、微信等的用户量迅速地上升，使人们主动在互联网上发表自己的意见；智能手机、平板电脑等新型移动设备的出现，也为人们随时随地在网上发表自己的观点提供了条件，这

个阶段数据产生的方式是主动的。

3. 感知式系统阶段

在经过前两个阶段之后，人类社会数据量第三次大的飞跃最终导致了大数据的产生，现在大家生活的这个时代正处于这个阶段。这个阶段数据产生的方式是自动的，随着科学技术的发展出现了诸如传感器之类的设备，并开始将这些设备广泛地布置于社会的各个角落，这些设备可以随时随地地产生数据，这个阶段促成了大数据的出现。这种数据产生的方式是自动的。

（二）大数据的产生原因

分析大数据的产生来源可以知道，大数据的产生原因主要是来自以下几个方面。

1. 数据存储成本的降低

大数据产生的重要前提是数据存储成本的大幅降低、存储硬件体积的日益减小。在早期，英特尔（Intel）创始人之一戈登·摩尔（Gordon Moore）提出著名的摩尔定律，在当时引起了广泛关注。摩尔定律是指，当价格保持固定时，大约每隔 18~24 个月，集成电路上元器件的数目便会增加一倍，其性能也将提升一倍，换句话说，每隔 18~24 个月，一美元所能买到的电脑性能至少翻一倍。

自摩尔定律产生至今已有半个世纪，半个多世纪以来，计算机硬件的发展规律基本符合摩尔定律，越来越接近定律之中的内容，慢慢地可以发现，硬件的处理速度、存储能力不断提升，与此相对应的是，硬件的价格在走下坡路。

除此之外，随着计算机硬件价格的逐步降低，计算机的体积也发生了变化，慢慢地变小，发展成今天可以随身携带的笔记本电脑。在前些年，英特尔公司发布了 14 纳米的晶体管，这比 21 纳米的晶体管体积缩小了 1/3，并且 14 纳米的晶体管在价格上更便宜、更节能、更环保。

由于存储器的价格下降速度飞快，人们才得以廉价保存海量的数据；由于存储器的体积越来越小，人们才可以便捷地携带海量的数据。这些都是大数据所带来的好处，这也在一定程度上促进了大数据时代的到来。

2. 生活的数字化驱动

物联网是新一代信息技术的重要组成部分，它的出现决定了物与物、人与物、人与人之间的互联。从本质上来说，因为物联网使得数据从主动式变为自动式，而大数据真正产生的原因正是由于人们生活中自动式数据的产生。

原本，人与机器之间、机器与机器之间的相互联系，就是要满足人与人之间的信息互动。在这种信息交互的过程中，因为需要对一些实际问题进行处理，所以逐步地由信息传送到信息感知再到面向分析处理的应用，人们在生活中所面对的各种各样的信息，都是需要传送到数据中心，经过数据中心的分析处理所得，得到结果之后，通过互联网等通信网络把这些信息传送到各处。对于互联网终端的设备，可以运用传感器等设备提取自己的有用信息，选择自己想要了解的数据。

现在，物联网应用在许多行业之中，如智能工业、智能农业、智能交通、智能电网、节能建筑、安全监控等。在应用的过程中需要借助于传感器，传感器等微小计算设备实现了无处不在的数据自动采集，这也意味着人们的数据收集能力的提高，为大数据的产生提

供了技术上的支持。

3. 社交网络的飞速发展

自早些年社交媒体相继问世,脸书(Facebook)、推特(Twitter)等各个软件受到大家的关注,这也意味着互联网时代的到来,即Web2.0时代的到来。进入Web2.0时代之后,人们对于互联网的运用越来越频繁,人们越来越多地利用互联网交流互动。而真正的数据爆发就产生于Web2.0时代,Web2.0时代最重要的标志就是用户原创内容。

在互联网时代,无论是机器、企业还是个人都需要获取数据,相对应地,也就产生新的数据。互联网的搜索巨头(谷歌)所处理的网页数量是无法计算的,每月所处理的数据超过400 PB,并在这个基础上逐步地增长;大家所熟悉的视频网站优酷网,每天都会上传7万小时的视频;购物的集中地淘宝网,在最开始的时候,其会员数量就已经达到3.7亿人,在线商品达到8.8亿件,每天的成交量更是超过数千万,单日数据产生量超过50 TB……以上全部的数据,其实都是海量数据的表现。

网络的普及与高速发展更为各种数据提供了集散场所,为数据生成的自动化、信息储存与传输的低成本化奠定了基础。或者可以这么认为,基础已经准备好了,大数据时代也就到来了。

伴随着信息技术的发展,社交网络也日趋成熟,传动的互联网也发生了改变,逐步地向移动互联网慢慢前进,在这些移动客户端中,计算机、手机以及平板电脑是屡见不鲜的,除此之外,还有很多更先进的传感设备或智能设备,在这种情况下所产生的数据量其实是无法计算的,其增长速度也是无法想象的,互联网时代所带来的数据流量正在迅猛地增长。

二、大数据的发展

(一) 大数据应用的发展

现在大数据正在发展运用,它存在于我们生活的方方面面,这给大数据问题的解决带来很多挑战。当然,大数据的广泛应用,使得很多学者和专家对大数据的研究越来越频繁。下面对大数据的发展中出现的一些大事件进行举例说明。

Hadoop 最开始是一个来自谷歌的编程模型包,其名称为 MapReduce。在设计之初,只是为了网页索引,被 Apache 软件基金会引入并成为分布式系统基础架构。在用户在没有清楚地知道布式底层细节的情况下,Hadoop 能够帮助用户开发分布式程序,充分发挥 Hadoop 集群的能力进行高速度的运算和存储,从而以一种可靠、高效、可伸缩的方式进行数据处理。Hadoop 框架最核心的设计就是 HDFS 和 MapReduce,HDFS 为海量的数据提供了存储,则 MapReduce 为海量的数据提供了计算。

起初,许多科学家和工程师都认为“大数据”特别可笑,觉得这只不过是一个营销术语。在前些年,“大数据”得到了部分美国知名计算机科学研究人员的认可,在业界范围内组织了计算社区联盟(Computing Community Consortium),并发表了题为《大数据计算》的白皮书,其肯定地表明了大数据的应用会带来更多机遇和挑战。

2009年,美国前总统贝拉克·奥巴马政府推出 data.gov 网站,作为政府开放数据计划的部分举措。这一网站自推出之后就得到广泛的使用,拥有超过4.45万的数据量集。

在这一类网站以及智能手机的应用程序中，可以追踪航班、召回一些具有质量问题的产品以及跟踪特定区域内失业率等信息，这一行动的推行给英国、肯尼亚等政府带来启示，他们相继推出了类似的措施。

2010年，肯尼斯·库克尔在《经济学人》上发表了长达14页的大数据专题报告《数据，无所不在的数据》。在报告中库克尔表示，世界上有着无法想象的巨量数字信息，并以极快的速度增长。从经济界到科学界，从政府部门到艺术领域，很多方面都已经感受到了这种巨量信息的影响。其实这也间接地表明，科学家和许多学者已经为这个现象作出了解释，命名为“大数据”。

2011年，如果你想要扫描2亿页的信息或4兆兆字节的磁盘存储，只需要几秒钟就可以完成，而不是像过去花费那么多的时间。与此同时，IBM的沃森计算机系统在智力竞赛节目《危险边缘》中打败了两名人类挑战者，《纽约时报》称这一刻为“大数据计算胜利”的时刻。

在同年，英国《自然》杂志曾出版专刊表明，如果人类能够很好地利用大数据，这将会给人类带来更多的成果，人类也会得到更多的机会去掌握科学技术，有利于社会的发展。

2012年，瑞士达沃斯召开的世界经济论坛的主题之一就是大数据，并且还发表了题为《大数据，大影响》的报告。报告中表示，现在数据已经成为一种新的经济资产类别，就像货币或黄金一样。

在同一年，美国政府报告中表示，每个联邦机构都要有一个与“大数据”有关的策略，响应此号召，奥巴马政府宣布了一项耗资两亿美元的大数据研究与发展项目。

软件公司Splunk在美国纳斯达克的成功上市，成为第一家上市的大数据处理公司。在之前，受经济危机的影响，美国经济持续地衰败，股市也受到巨大的波动，Splunk在上市后的第一天就表现得非常出色，令大家印象深刻。

后来美国国务卿希拉里·克林顿宣布了一个名为“数据2X”的公私合营企业，用来收集统计世界各地的妇女和女童在经济、政治和社会地位方面的信息。

2014年，世界经济论坛以“大数据的回报与风险”为主题发布了《全球信息技术报告（第13版）》。报告的观点是，在以后的几年里，面对各种信息通信技术的政策将会变得越来越重要。与大数据有关的产业开始变得越来越活跃，技术在逐步地发展，创新也在逐步地推广中，这使得许多国家的政府都意识到大数据的重要性，大数据有利于经济的发展、还能够促使人们生活水平的提高，更加重要的是大数据还可以保障国家安全。在2014年5月，美国白宫发布了2014年全球“大数据”白皮书的研究报告《大数据：抓住机遇、守护价值》。在报告中，美国鼓励各个部门积极地使用大数据，以此来推动社会的进步，建立市场与大数据相关联的机构；与此同时，还需要相应的框架、结构与研究，通过这种方式来保护美国人的个人隐私，保证公平。

现在，以互联网巨头推动为主的大数据应用，无论是对促进经济社会发展，还是对保障国家安全，都扮演着十分重要的角色。然而，作为国家经济基石的传统产业仍基本游离在大数据应用之外，这表明大数据应用之路才刚刚开始。

在当今社会，大数据如影随形，它是一场革命，能够改变人类的生活、工作以及思维方式。如此巨大的数据来源，已经引起了学术界、企业界和政府的广泛关注，他们致力于

大数据更新的研究。

(二) 大数据技术的发展

大数据技术是一种新一代的技术和构架，它的成本比较低，能够快速地采集、处理和分析技术，在数以万计的数据中选择有价值的信息。大数据技术的发展以及广泛应用，让大家能够更简单、更方便、更快捷地处理数据，并且在大数据的影响下，能够改变许多行业的商业模式。因此，可以将大数据技术的发展分为以下几个方向。

1. 大数据采集与预处理方向

这个方向有一个普遍性的问题，那就是数据的多源和多样性，造成的后果就是数据的质量存在差异，这就会给数据的可用性带来问题。根据这些问题，我们可以发现，现在许多的公司都已经推出了多种数据清洗和质量控制工具，如 IBM 的 Data Stage。

2. 大数据存储与管理方向

大数据的数量是无法想象的，这给存储和管理带来了问题。存储规模的巨大、管理的复杂，就需要研究者想方设法地解决问题，而分布式文件系统和分布式数据库相关技术的发展就可以处理这些问题。面对大数据存储和管理方向，需要特别注意的是，大数据索引和查询技术、实时及流式大数据存储与处理的发展。

3. 大数据计算模式方向

因为大数据处理多样性的要求，现在已经出现了多种典型的计算模式，这些计算模式涉及的内容众多，例如，大数据查询分析计算或批处理计算，还有流式计算或迭代计算，图计算和内存计算，当然也可以将这些计算模式结合起来，这对大数据处理具有重要的意义。

4. 大数据分析挖掘方向

现在大数据的应用越来越广泛，数据也随之迅速地增长，在对数据进行应用的同时，还需要对其进行分析与挖掘。现在大数据也能进行自动化分析，并且对自动化分析的要求越来越高，工程师们制造了许多大数据分析工具和产品，例如基于 MapReduce 开发的数据挖掘算法。

5. 大数据可视化分析方向

对众多繁杂的数据信息进行可视化分析，能够帮助决策者发现数据挖掘的价值，还可以推动大数据的发展。针对可视化分析，现在许多公司都在进行研究，致力于将可视化分析与数据结合在一起，众多相关的产品也会应运而生。其中，有一个可视化的工具 Tableau 成功上市，由此可以看出大数据可视化的重要性。

6. 大数据安全方向

现在，生活的方方面面都需要大数据，大数据在迅速地增长，与此相对应的是，这也带来了数据的安全问题。一方面因为大数据众多，更容易在网络上被发现；另一方面大数据拥有一些有价值的数据，会吸引攻击者破坏数据信息。当我们利用大数据进行数据分析与挖掘来获取商业价值时，这也可能会引起黑客的注意，从而攻击我们，截取信息。所以，大数据的安全应该是大家非常关心的话题。可以利用文件访问控制去限制呈现对数据的操作、基础设备加密、匿名化技术和加密保护技术等，都在最大限度地保护数据安全。

1.2 大数据的概念、特征与类型

一、大数据的概念

大数据是一个比较宽泛的概念，如果只是从最浅析的意思来理解，可能就是信息繁多、规模庞大。然而，如果只是从数量上的庞大分析，是无法看出大数据和以往的“海量数据”有什么不同。目前学者和专家们并没有对大数据给出一个统一的概念，下面我们来分析这些理解不同。

百度百科中对大数据的定义是，大数据也可以称为巨量资料，它主要指的是包含的资料是无法想象的，也不能利用现在的软件去得出结果，在一定的时间内，可以帮助企业管理、处理信息。

维基百科中对大数据的定义是，大数据的信息量是巨大的，人们根本不能利用某些工具在规定时间内去处理、管理并且把各种各样的数据整理成人们所需要的信息。其实说的就是，大数据是一个极其大的数据集，这种数据集还不能采用传统的方法进行处理。

根据美国国家标准与技术研究院（National Institute of Standards and Technology, NIST）发布的研究报告的定义，大数据是现在人们常用的词语，在网络时代、数字时代、信息化时代的今天，大数据这个词语无处不在，在我们生活的许多方面都可以看到大数据的影子。现在的海量数据使之前无法解决的问题，都能够慢慢地得到解决。

研究机构 Gartner Group 也给出了自己的理解：运用传统工具无法处理的信息，可以采用大数据完成，也就是大数据需要新的处理模式，才可以富有强大的决策能力、发现力以及优化各个流程的能力，才能获取多样化的信息。

麦肯锡公司给出的定义是，传统的数据库对于信息的获取、存储、处理和分析的能力，已经无法满足信息量日益发展的今天，使用大数据却可以摆脱传统的数据形式，拥有超大的信息，规模庞大到无法想象。这个定义被界定成主观性的，其实这说的是，并不是大于一个特定数字的 TB 才叫作大数据。伴随着科学技术的迅速发展，对于大数据的信息需求也会随之产生变化；并且这个定义在不同的行业也会发生一定的变化，这取决于一个特定行业通常使用何种软件和数据集有多大。所以，在今天不同的行业中，大数据的量是不同的，其范围可以从几十 TB 到几 PB。

大数据的概念特别广泛，每个人对大数据都有自己的理解。在前面的观点中加上作者自己的理解，给出的解释是：大数据是在体量和类别特别大的杂乱数据集中，发掘隐藏在更深处的有价值的信息。一方面需要注重大数据数量之大，“大”其实只是一个表面的含义；另一方面，需要更加注重对“数据”的分析和应用，将有价值的信息充分应用，这才是学者们应该着重关注的点。

大数据其实是一个从量变到质变的转化过程，它代表着在现实生活中，无论是在经济方面，还是在社会实践方面，数据作为一种资源都发挥着重要的作用，与之有关的技术、产业、应用都会相互影响、共同前进。从技术角度进行理解，大数据形成质变以后会出现

一些新的问题,也就是数据从静态变为动态,从简单的多维度变成巨量维度,并且其种类变得越来越多,现在的分析方法和技術已经不能满足这种数据应用。这些数据的采集、分析、处理、存储、展现都涉及高维复杂多模态计算过程,涉及异构媒体的统一语义描述、数据模型、大容量存储建设,涉及多维度数据的特征关联与模拟展现。但是,从最本质的角度进行分析,大数据还是为了其应用价值,如果大数据没有价值,那也没有什么意思了。

二、大数据的特征

现在,大家比较一致的理解就是互联网数据中心对大数据的定义,从这个定义中可以分析出大数据的基本特征是规模性 (Volume)、多样性 (Variety)、高速性 (Velocity)、价值性 (Value),也就是经常说的 4V 特性。

(一) 规模性

根据 TechWeb 的报道,在一天之内,互联网上所产生的全部信息量,能够刻 1.68 亿张光盘;可以发出 2940 亿封的邮件;发出的社区帖子能够达到 200 万个,这个数字是美国《时代》杂志 770 年的文字总量;如果是 1.72 亿人在登录 Facebook,一共需要的时间是 47 亿分钟,在上面要传送 2.5 亿张图片,把这些图片全部打印出来,差不多有 80 座巴黎埃菲尔铁塔的高度。随后,全球数据量已经从 TB ($1024\text{ GB}=1\text{ TB}$) 级别跃升到 PB ($1024\text{ TB}=1\text{ PB}$)、EB ($1024\text{ PB}=1\text{ EB}$) 乃至 ZB ($1024\text{ EB}=1\text{ ZB}$) 级别。当然,数据量的变化只是最初阶段,更大的规模还在后面。IDC 预测,未来几年,全球数据量每隔两年翻一番,2020 年将达到 35 ZB。

(二) 多样性

现代社会的传感器、智能设备还有其他的社会技术,都在与日俱增,不知道何时就出现了新的技术。在这种大环境下,企业的数据也就变得日益复杂,因为这牵涉的内容众多,不仅仅是传统的关系型数据,还有网页、搜索索引、电子邮件等原始、半结构化和非结构化数据。

所有的数据类型是可以种类表示的。在这之中,爆发式增长的一些数据,如互联网上的文本数据、位置信息、传感器数据、视频等,如果是采用企业中主流的关系型数据,是很难进行操作的,它们都是非结构化数据中的一种。

自然,这些数据并非是全新的,有一些是从过去就保留下来的,有所不同的是,不仅仅是需要对这些数据进行存储,还需要分析这些数据,从所有的信息内容中获取有价值的信息,如监控摄像机中的视频数据。现在,许多的企业都设置了监控摄像机,如超市、便利店等,起初是为了防范盗窃,但是现在企业利用监控摄像机的数据分析顾客的购买行为。

例如,美国高级文具制造商万宝龙,他们过去对顾客的分析都是根据经验和直觉来判断,以此决定商品如何布局,现在他们利用监控摄像头分析顾客在店内的消费行为,更好地对商品排列布局,吸引消费者。通过分析监控摄像机的数据,将最想卖出去的商品移动到最容易吸引顾客目光的位置,使得销售额提高了 20%。

（三）高速性

在信息时代，大数据的一个显著特征是数据产生和更新的速度，这个速度是无法形容的。就像搜集和存储的数据量和种类发生了变化一样，生成和需要处理数据的速度也在变化。不能把速度的定义，简单地认为是与数据存储相关的增长速率，应该动态地把这个概念应用到数据，即数据流动的速度。

现在，伴随着科学技术的发展，对数据智能化和实时性的要求越来越高，例如，在外出开车的时候，需要借助于导航仪查询路线；在吃饭之前通常需要先对餐厅作一个了解；很多人看到美食都会在微博、微信上展示……这些人与人之间、人与机器之间的信息交流互动，都会带来数据交换，在数据交换过程中最重要的一步是降低延迟，实时地呈现给用户。

（四）价值性

一般来说，大量的非传统数据中会含有一些很重要的内容，不过很麻烦的是，如何在万千的信息中选择出有意义的内容，之后提取这些数据进行分析，从中得到有价值的信息内容，然而有价值的内容信息也只是很少的一部分。这就相当于沙里淘金，在大堆沙子中只存在一点点金子，只有那沙堆中的一点点金子是有价值的。现在，监控视频运用得越来越多，许多公共场所都装有监控设备，如银行、地铁等地点，并且这些地点的摄像头是24小时的，时刻运转，产生的视频数据也是很大的。通常来说，这些视频数据基本上是没有作用的，大家对此不会过多的关注，但是在某些特殊情况下，如公安部门需要获取犯罪嫌疑人的体貌特征，虽然有效的视频信息很短，但是却给公安人员带来极大的帮助。因为监控视频中不知道哪几秒是对大家有用的，所以需要全部保留下来，在以后可能就会发挥很大的作用。

然而，在研究人类行为的社会学家眼中，这些监控视频数据是很重要的资料，他们对其非常重视，从视频中某些表现可以发现人类的行为特征。因此，大数据的价值密度低是指相对于特定的应用，信息有效与否是相对的，数据的价值也是相对的，对于某一个应用，一些数据可能是没用的，但是在另一领域中，这些数据却是极其重要的。换一种思维理解，可以把这些数据重新组合和处理，之前没有发现的价值也是很难猜测的。

大数据与传统数据的概念是不同的，最明显的区别就在于大数据的4V特性。之前的“海量数据”概念只强调量，而大数据不仅仅是描述数据的量，还表现数据的规模、高速性以及复杂的形式，通过专业化的处理来获取有价值的信息。

三、大数据的类型

大数据不仅仅是数量众多、规模巨大，还表现在数据的类型上，在大量的信息内容中，只有大约20%的数据隶属于结构化数据，大约80%的数据隶属于分布在社交网络、物联网、电子商务等领域的非结构化数据。因为现在所创造的技术产生的数据是当前的方法所处理不了的，而机器数据越来越重要，数据将会成为一种自然资源。