分析工具先进

案例经典实用

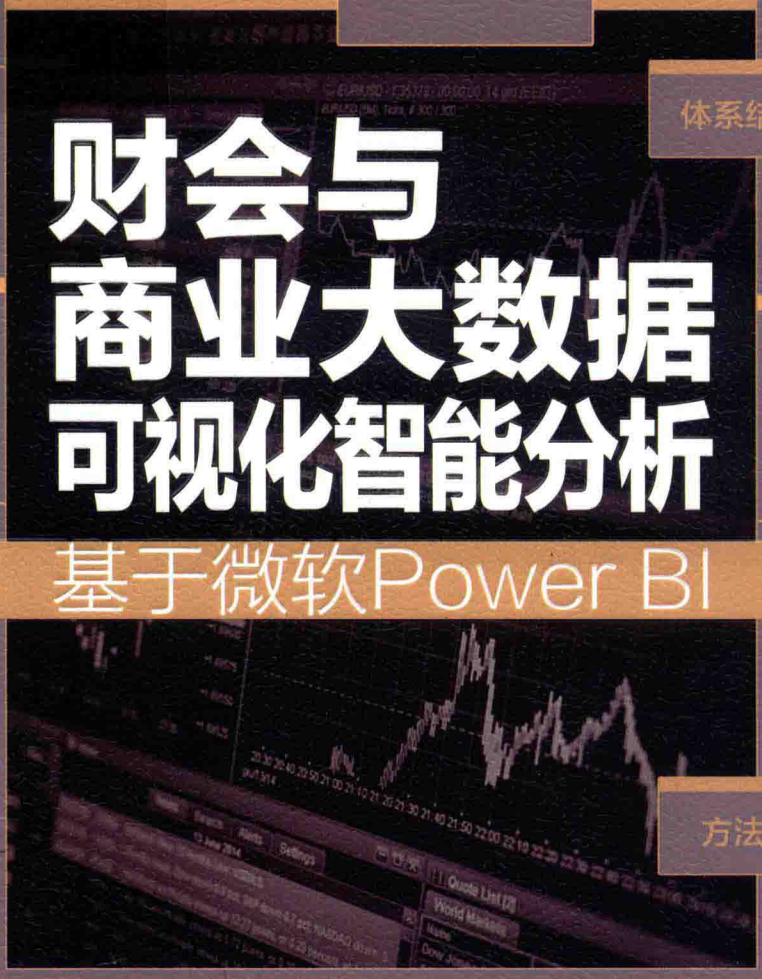
体系结构完整

财会与商业大数据可视化智能分析

基于微软Power BI

财会转型容易

配套资源丰富



方法思路独特

汪刚◎编著

1 **手把手**教您从数据可视化入手学习数据分析

2 **透彻讲解**财会与商业大数据可视化智能分析的**思路和方法**

3 提供**113**个可操作**案例**，**53**个相关**数据文件**

4 提供案例**讲解视频**
一步步教您**实战演练**，为**自学**提供极大便利

清华大学出版社

财会与商业大数据可视化智能分析

——基于微软Power BI

汪　刚　编著

清华大学出版社
北　京

内 容 简 介

随着“大智移云物”技术的发展，很多财会人员正积极向智能可视化的财务分析方向转型。2019年2月，国际著名咨询机构Gartner公司发布的《商业智能和分析平台魔力象限》年度报告显示，微软超越一切对手，再次成为最具领导力和超前愿景的BI公司。本书以微软Power BI为工具，以案例驱动方式讲解数据分析(数据获取与整理、数据建模、数据可视化)的一般思路及方法。

本书共分9章。第1章介绍了大数据、云计算、商业智能及数据分析思路等内容；第2章介绍了Power BI的特点、系列组件、安装及账号注册等内容；第3章以一家连锁烘焙店的刷卡记录为案例数据帮助读者快速体会Power BI可视化智能分析的基本应用流程；第4～6章系统地介绍了数据整理、数据建模、数据可视化的数据分析及展现过程；第7章讲解了Power BI在线应用等内容；第8章和第9章以两个完整案例讲解了财会与商业大数据可视化智能分析的思路和方法。

全书共设计了113个可操作的案例，涉及53个原始数据文件和结果数据文件等。

本书适用于企事业单位从事数据分析工作的人员，也可以作为高校经管类相关专业老师和学生进行大数据可视化分析的参考书籍。

图书在版编目(CIP)数据

财会与商业大数据可视化智能分析：基于微软Power BI / 汪刚 编著. －北京：清华大学出版社，2019
ISBN 978-7-302-53641-3

Ⅰ. ①财… Ⅱ. ①汪… Ⅲ. ①可视化软件－应用－财务会计－数据处理 ②可视化软件－商业信息－数据处理 Ⅳ. ①F234.4-39 ②F713.51-39

中国版本图书馆 CIP 数据核字(2019)第 186761 号

责任编辑：刘金喜
封面设计：周晓亮
版式设计：孔祥峰
责任校对：牛艳敏
责任印制：刘海龙

出版发行：清华大学出版社
网　　址：http：//www.tup.com.cn，http：//www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　邮　　编：100084
社 总 机：010-62770175　　邮　　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：北京嘉实印刷有限公司
经　　销：全国新华书店
开　　本：190mm×260mm　　印　　张：16.25　　字　　数：375 千字
版　　次：2019 年 9 月第 1 版　　印　　次：2019 年 9 月第 1 次印刷
定　　价：58.00 元

产品编号：083798-01

前言

财务的未来是信息化、自动化、数字化和智能化。“大智移云物”——大数据、人工智能、移动互联网、云计算和物联网等技术的快速发展，正在促使未来成为一个“万物互联、无处不在、虚实结合、智能计算、开放共享”的智能时代，数据成为企业的核心资产，财务部门应积极尝试新兴技术，能够更广泛、更智能地收集数据、加工数据和分析数据，实现财务数字化转型，帮助企业提升经营能力、洞察商机并预测未来。

【财会人员面临转型】

随着“大智移云物”技术的发展、财务机器人(Robotic Process Automation，RPA)的出现，传统标准化的财务会计工作正在逐渐被RPA替代，如何转型是每一个财会人员需要思索的问题。

从算盘到计算器再到计算机，财务工具的不断变革大幅改进了财务工作的效率。德勤公司在2018年的《关键时刻——数字化世界中的财务》报告中认为，云计算、流程机器人、可视化、高级分析、认知计算、内存计算和区块链七项技术对财务的影响愈发显著，它们共同构筑了新时代下的财务工具集。

可视化(Visualization)是利用计算机图形学和图像处理技术，将数据转换成图形或图像在屏幕上显示出来，再进行交互处理的理论、方法和技术。可视化的运用有如下意义：提供实时信息，大大加快数据处理速度，使庞大的数据得到有效利用；多维显示数据，展现全貌，更好地展现各个因素之间的关联；简化复杂性，更加直观地展示复杂信息；增强理解，便于对话、探索和交流。可视化具有广泛的应用领域，包括科学、教育、信息、数据、地理、医学影像、产品、软件、工程制图和立体渲染等。可视化在财务中主要应用于数据统计与分析，能更清楚、直观和多维度地传达信息及展示趋势，帮助企业“看到”涉及重大决策和发展的事项。

可视化是数据分析的重要内容。目前进行数据分析的工具有很多，包括：偏重于数据采集的Kettle、Python；偏重于数据仓库的Hadoop、Teradata等；偏重于数据挖掘的SaS、SPSS、R、Python等；偏重于数据可视化的BIEE、Tableau、Power BI、帆软等。这些工具有的需要具备计算机编程知识，有的需要具备统计学知识，因此都不适合财经类人员快速掌握数据分析技能。对于熟悉业务、财务的财会人员来说，从数据可视化入手学习数据分析是最佳的选择。

【为什么选择微软Power BI】

目前，市场上用于数据可视化的工具较多，有BIEE、Tableau、Power BI、SAP BI、帆软等。本书以微软Power BI为工具，讲解数据分析(数据获取与整理、数据建模、数据可视化)的一般应用。

Power BI是微软官方推出的一个让非数据分析人员也能做到有效地整合企业数据，并快速准确地提供商业智能分析的数据可视化神器和自助式BI分析工具。Power BI既是员工的个人报表和数据可视化工具，还可用作项目组、部门或整个企业背后的分析和决策引擎。2019年2月，国际著名咨询机构Gartner公司发布的《商业智能和分析平台魔力象限》年度报告显示，微软超越一切对手，再次成为最具领导力和超前愿景的BI公司。Gartner公司对微软的评价是："微软是领导者。他具有全面而富有远见的产品路线图，致力于打造覆盖所有分析场景的全局统一、人人可用的Power BI。"

【本书体系结构】

本书共分9章，具体如下。

第1章"商业智能与数据分析概述"从理论上介绍了大数据、云计算、商业智能及数据分析思路等内容。

第2章"微软Power BI概况"介绍了Power BI的特点、地位、应用模式、系列组件、安装与账号注册、界面等内容。

第3章"快速实践Power BI"以一家连锁烘焙店为案例让学员快速体会Power BI可视化智能分析的基本应用。

第4章"数据整理"主要讲解如何进行数据获取与数据处理。

第5章"数据建模"主要讲解在Power BI中如何创建关系、度量值，如何使用DAX——数据分析表达式。

第6章"数据可视化"讲解了数据可视化的原则、常用可视化图表、自定义可视化图表等内容。

第7章"Power BI在线服务"讲解了在线发布、创建仪表板、分享与协作、移动应用等内容。

第8章"案例1——海信电器财务数据可视化智能分析"以海信电器公司5年的财务数据进行财务报表可视化和财务分析可视化。

第9章"案例2——某连锁店大数据可视化智能分析"以某运动品牌连锁店业务数据(模拟)进行产品分析、区域分析等可视化分析。

【本书案例资源】

- 本书共设计了113个可操作的案例，涉及原始数据文件和结果数据文件53个。
- 本书提供各案例操作视频。

上述案例资源可通过扫描下方二维码直接用手机下载，或将链接推送到邮箱，在PC端下载。


数据文件下载


案例视频下载

【致谢】

感谢第二届微软Power BI可视化大赛冠军刘成城与我分享大赛和财务数据分析可视化的经验，以及提供了相关案例数据供我学习。

网易云课堂的赵文超和BI佐罗老师在基于企业实践经验上带来了精彩纷呈的Power BI课程，并在我制作案例数据遇到困难时给予我耐心的解答，在此表示感谢。

限于作者水平，对于书中的疏忽及错漏之处，诚挚地希望广大读者给予批评指正。

服务邮箱：476371891@qq.com。

作 者

2019年6月

目 录

第 1 章

商业智能与数据分析概述

引言

现代管理之父彼得·德鲁克曾经说：“知识已经成为关键的经济资源和竞争优势的主要来源。”如何将数据转化为知识？这就需要商业智能(Business Intelligence，BI，也叫商务智能)发挥作用。商业智能是将不可用数据转变为可行的见解的过程。当正确实施后，商业智能将提高可见性、对客户行为的洞察力及效率等。

新经济时代的赢家是把顾客、供应商等相关的运营数据整合、分析和共享，转化为信息，并进一步分析得到知识，提高企业商业智能从而保持盈利的企业。面对激烈的竞争，传统的决策支持系统(Decision Support System，DSS)已难以支撑，而作为后ERP时代的信息化应用，商业智能恰好为企业提供了这样的一种利器。

本章将从大数据与云计算、商业智能概述、商业大数据分析思路3个方面让大家初步认识商业智能与数据分析。

本章知识结构的思维导图，如图1-1所示。

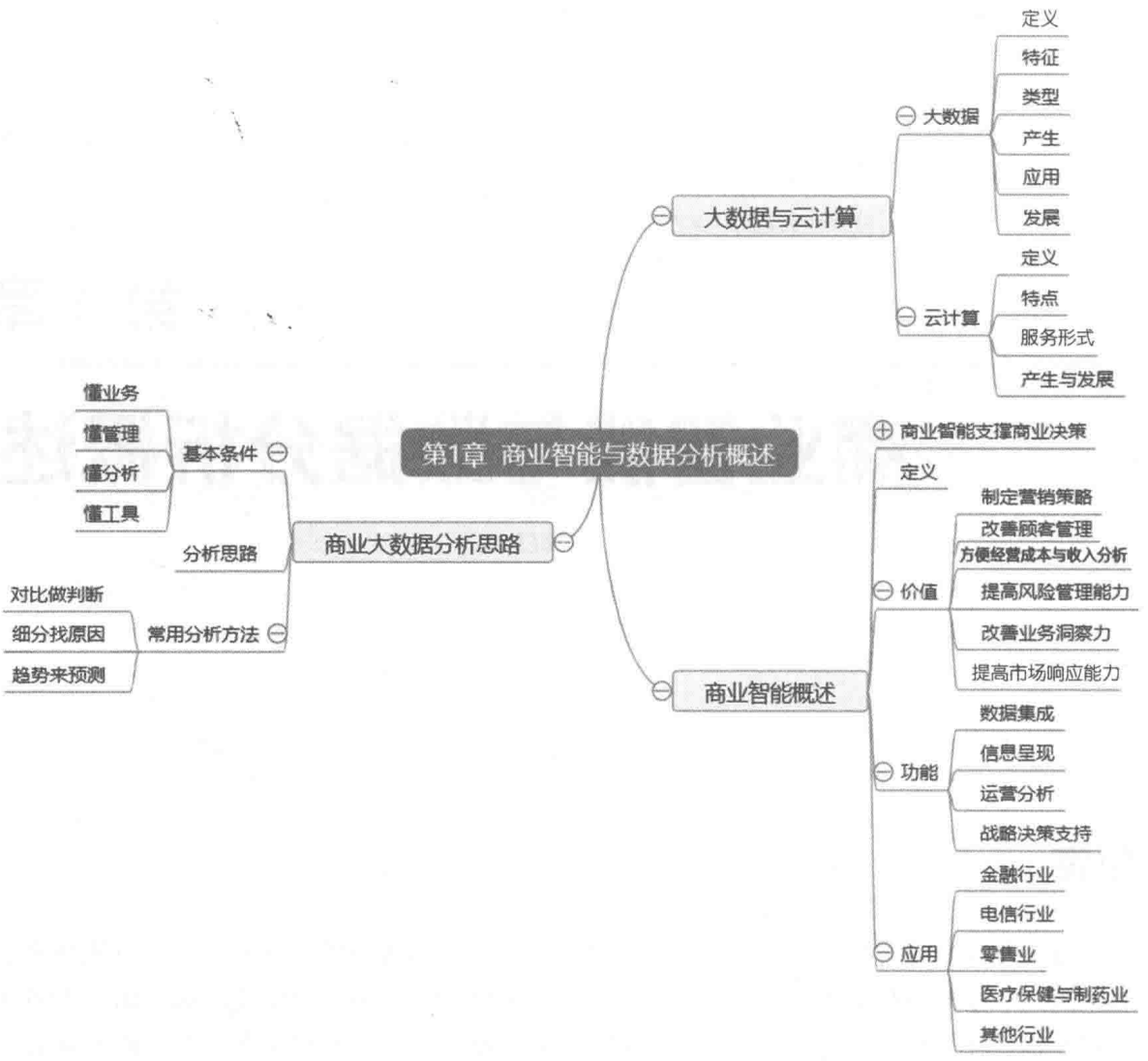


图 1-1　本章知识结构的思维导图

1.1 大数据与云计算

1.1.1 大数据

1. 大数据的定义

大数据本身是一个比较抽象的概念，单从字面来看，它表示数据规模的庞大。但是仅仅数量上的庞大显然无法看出大数据这一概念和以往的海量数据(Massive Data)、超大规模数据(Very Large Data)等概念之间有何区别。针对大数据，目前存在多种不同的理解和定义。

维基百科对“大数据”(Big Data)的解读是：大数据，或称巨量数据、海量数据、大资料，指的是所涉及的数据量规模巨大到无法通过人工在合理时间内达到截取、管理、处理，并整理成为人类所能解读的信息。

百度百科对“大数据”的定义为：大数据，或称巨量资料，指的是所涉及的资料量规模巨大到无法透过目前主流软件工具在合理时间内达到撷取、管理、处理，并整理成为帮助企业经营决策更积极目的的资讯。

麦肯锡在其报告《Big data：The nextfrontier for innovation，competition and productivity》中给出的大数据定义是：大数据指的是大小超出常规的数据库工具获取、存储、管理和分析能力的数据集。但它同时强调，并不是说一定要超过特定TB值的数据集才能算是大数据。

研究机构Gartner认为：“大数据”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的海量、高增长率和多样化的信息资产。从数据的类别上看，“大数据”指的是无法使用传统流程、工具处理或分析的信息。它定义了超出正常处理范围和大小、迫使用户采用非传统处理方法的数据集。

大数据代表着数据从量到质的变化过程，代表着数据作为一种资源在经济与社会实践中扮演着越来越重要的角色，相关的技术、产业、应用、政策等环境会与之互相影响、互为促进。从技术角度来看，这种数据规模质变后带来新的问题，即数据从静态变为动态，从简单的多维度变成巨量维度，而且其种类日益丰富，超出当前分析方法与技术能够处理的范畴。这些数据的采集、分析、处理、存储、展现都涉及复杂的多模态高维计算过程，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。然而，大数据发展的最终目标还是挖掘其应用价值，没有价值或者没有发现其价值的大数据从某种意义上讲是一种冗余和负担。

2. 大数据的特征

由维克托·迈尔-舍恩伯格和肯尼思·库克耶编写的《大数据时代》中提出大数据具有4V特征：规模性(Volume)、高速性(Velocity)、多样性(Variety)和价值性(Value)，如图1-2所示。

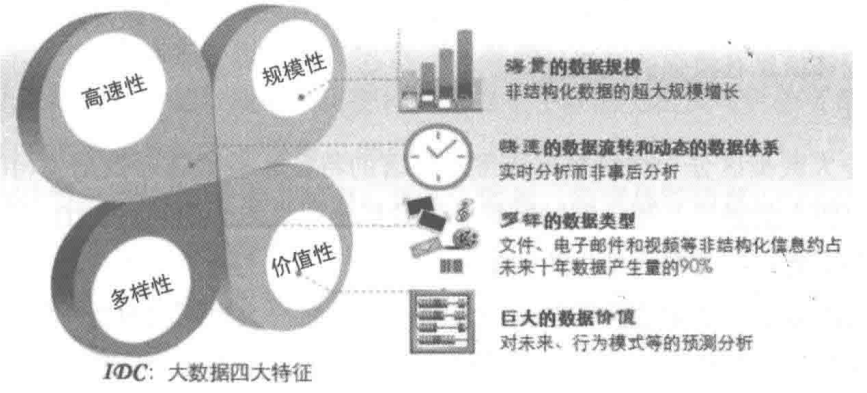


图 1-2　大数据的 4V 特征

1) 规模性

随着信息化技术的高速发展，数据开始爆发性增长。大数据中的数据不再以几个GB或几个TB为单位来衡量，而是以PB(1000个T)、EB(100万个T)或ZB(10亿个T)为计量单位。2010年，麦肯锡全球研究院估计，全球企业在硬盘上存储了超过7EB(1EB等于10亿GB)的新数据，消费者在PC和笔记本电脑等设备上存储了超过6EB的新数据，数据总量相当于美国国会图书馆中存储的数据的5.2万倍。据统计，目前整个人类社会总共拍摄了超过3.5万亿张照片，绝大多数是数码存储的照片。如今人们每两分钟拍摄的照片数就比整个19世纪拍摄的照片总数还要多。Facebook已经成为世界上最大的照片库，目前全球累计已有超过1400亿张照片发布在 Facebook网站上。

社交网络、移动网络及各种智能终端等，都可成为数据的来源。例如，淘宝网近4亿的会员每天产生的商品交易数据约为20TB；Facebook约10亿的用户每天产生的数据超过300TB；Google每天通过云计算平台处理的数据超13.4PB。因此，我们迫切需要智能的算法、强大的数据处理平台和新的数据处理技术来统计、分析、预测和实时处理如此大规模的数据。

2) 多样性

多样性主要体现在数据来源多、数据类型多和数据之间关联性强3个方面。

(1) 数据来源多。企业所面对的传统数据主要是交易数据，而互联网和物联网的发展，带来了诸如社交网站、传感器等多种来源的数据。

而由于数据来源于不同的应用系统和设备，决定了大数据形式的多样性，大体可以分为三类：一是结构化数据，如财务系统数据、信息管理系统数据、医疗系统数据等，其特点是数据间因果关系强；二是非结构化数据，如视频、图片、音频等，其特点是数据间没有因果关系；三是半结构化数据，如HTML文档、邮件、网页等，其特点是数据间的因果关系弱。

(2) 数据类型多，并且以非结构化数据为主。传统的企业中，数据都是以表格的形式保存。而大数据中有70%～85%的数据是非结构化和半结构化的数据，如图片、音频、视频、网络日志、链接信息等。

(3) 数据之间关联性强，频繁交互，如游客在旅游途中上传的照片和日志，就与游客的位置、行程等信息有很强的关联性。

3) 高速性

高速性是大数据区分于传统数据挖掘最显著的特征。根据国际数据公司(IDC)的一份名为“数字宇宙”的报告，预计到2020年全球数据使用量将会达到35.2ZB。在如此海量的数据面前，处理数据的效率就是企业的生命。

大数据与海量数据的重要区别在两个方面：一方面，大数据的数据规模更大；另一方面，大数据对处理数据的响应速度有更严格的要求，实时分析而非批量分析，数据输入、处理与丢弃立刻见效，几乎无延迟。数据的增长速度和处理速度是大数据高速性的重要

体现。

既有的技术架构和路线，已经无法高效处理如此海量的数据，而对于相关组织来说，如果投入巨大的采集信息则无法通过及时处理反馈有效信息，那将是得不偿失的。可以说，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。

4) 价值性

尽管我们拥有大量数据，但是发挥价值的仅是其中非常小的部分。大数据背后潜藏的价值巨大，如美国社交网站 Facebook有10亿用户，网站对这些用户信息进行分析后，广告商可根据结果精准投放广告。对广告商而言，10亿用户的数据价值上千亿美元。据资料报道，2012年，运用大数据的世界贸易额已达60亿美元。

由于大数据中有价值的数据所占比例很小，而大数据真正的价值体现在从大量不相关的各种类型的数据中挖掘出对未来趋势与模式预测分析有价值的数据，并通过机器学习方法、人工智能方法或数据挖掘方法深度分析，运用于农业、金融、医疗等各个领域，以创造更大的价值。

3. 大数据的类型

大数据不仅体现在数量大，也体现在数据类型多，如此海量的数据，仅有20%左右属于结构化数据，80%的数据属于广泛存在于社交网络、物联网、电子商务等领域的非结构化数据。由于我们创造的技术产生的数据已经远远超越了目前的方法和工具所能处理的范畴，而机器数据越来越重要，因此，数据将会成为一种自然资源。

1) 按照数据结构分类

按照数据结构，可将数据分为结构化、非结构化、半结构化。

(1) 结构化数据。结构化数据是存储在数据库中，并可以用二维表结构来逻辑表达实现的数据。结构化数据指的是关系模型数据，即以关系型数据库表形式管理的数据。绝大多数的企业业务数据都以此格式进行数据存放。

(2) 非结构化数据。相对于结构化数据而言，不方便用数据库二维逻辑表来表现的数据即称为非结构化数据，包括所有格式的办公文档、文本、图片标准通用标记语言下的子集XML、HTML、各类报表、图像和音频/视频信息等。

非结构化数据库是指其字段长度可变，并且每个字段的记录又可以由可重复或不可重复的子字段构成的数据库，它不仅可以处理结构化数据(如数字、符号等信息)，而且更适合处理非结构化数据(如全文文本、图像、声音、影视、超媒体等信息)。

非结构化Web数据库主要是针对非结构化数据产生的，与以往流行的关系数据库相比，其最大的区别在于它突破了关系数据库结构定义不易改变和数据固定长度的限制，支持重复字段、子字段及变长字段，并实现了对变长数据和重复字段进行处理和数据项的变长存储管理，在处理连续信息(包括全文信息)和非结构化信息(包括各种多媒体信息)中有

着传统关系型数据库所无法比拟的优势。

(3) 半结构化数据。所谓半结构化数据，就是介于完全结构化数据(如关系型数据库、面向对象数据库中的数据)和完全无结构的数据(如声音、图像文件等)之间的数据，如HTML文档就属于半结构化数据。它一般是自描述的，数据的结构和内容混在一起，没有明显的区分。

2) 按照产生主体分类

按照产生主体，可将数据分为企业数据、机器数据、社会化数据。企业数据包括CRM系统中的消费者数据、传统的ERP数据、库存数据及账目数据等；机器数据包括呼叫记录、智能仪表、工业设备传感器、设备日志、交易数据等；社会化数据包括用户行为记录、反馈数据等，如Twitter、Facebook等社交媒体平台。

(1) 企业数据(Enterprise Data)。2010年，全球企业新存储的数据超过了7000PB，全球消费者新存储的数据约为6000PB，每一天都有无数的数据被收集、交换、分析和整合。数据已经如一股“洪流”注入了世界经济，成为全球各个经济领域的重要组成部分。数据将与企业的固定资产、人力资源一样，成为生产过程中的基本要素。

2011年，麦肯锡在其研究报告《大数据：下一个创新、竞争和生产率的前沿》中指出，在美国，仅制造行业就拥有比政府还多一倍的数据，此外，新闻业、银行业、医疗业、投资业、零售业都拥有可以与政府相提并论的海量数据。

据IDC发布的《中国大数据技术与服务市场2012—2016年预测与分析》报告显示，该市场规模将会从2011年的7 760万美元增长到2016年的6.17亿美元，未来5年的复合增长率达51.4%，市场规模增长近7倍。庞大的数据来源所带来的量化转变在企业界已经迅速蔓延。

(2) 机器数据(IT Data)。大数据中，机器数据是份额最大且增长最快的一部分。每个现代企业机构，无论规模大小，都会产生海量的机器数据，如何管理和利用机器数据，进行业务创新并获取竞争优势，已经成为目前企业或机构所面临的关键任务。

机器数据，顾名思义，是由机器(软硬件系统)产生的数据，也是大数据最原始的数据类型，它通常包括所有软硬件设备生产的信息，如日志文件、交易记录、网络消息、传感器采集的数据等，这些信息几乎包含了所有客户、交易、设备等元素的动作行为。

在大数据时代，结合IT运维、系统安全、搜索引擎、电子商务等特定应用的需求，实现大数据环境下机器数据的存储、管理、检索和分析，将是目前企业或机构管理和利用机器数据的重点所在。

(3) 社会化数据(Social Data)。随着社交网络的流行，国内外社会化媒体得到了迅猛发展。截止到2012年10月，Facebook的用户数超过10亿，Twitter的用户数超过5亿。据中国互联网络信息中心最新发布的报告显示，中国的网民已达5.55亿，其中超过4亿的用户分布在微博、SNS、个人空间等社会化媒体上。

集中在社会化媒体上的庞大用户群及发生的用户行为将会产生巨量的数据回馈，包括

评论、视频、照片、地理位置、个人资料、社交关系等。由用户在社会化媒体中产生或分享的各类信息即为社会化数据。

社会化数据与以前采集的静态的、事务性数据完全不一样，它具有实时性和流动性。人们在社会化媒体上通过交流、购买、出售和其他日常生活活动以免费的方式提供着大量信息。这些数据由每个网民的微行为汇集而成，蕴含着巨大的价值，将带来政府在公共管理方面、企业在市场调研和营销方面的变革。

3) 按照数据作用方式分类

按照数据作用方式，可将数据分为交易数据、交互数据和传感数据。

(1) 交易数据。交易数据即为ERP、电子商务、POS机等交易工具带来的交易数据。在实际应用中，组织数据与互联网数据尚未有效整合，在数据处理中，杂乱的、海量的、沉睡的数据严重地影响了数据的有效利用。面对这些挑战，人们急需综合的大数据平台、快速有效的算法来统计、分析和预测组织产生的交易数据，以便更好地为决策进行服务。

(2) 交互数据。交互数据即为微信、微博、即时通信等社交媒体带来的数据。社群网站的盛行，带动了以非结构化数据为主的大数据分析，促使企业不再只满足于点状的交易数据，例如，产品卖掉了、顾客突然解约等属于点状的交易数据，而我们必须将点状的交易数据转为探究线状的互动数据，如为什么这项产品卖掉了、顾客为什么突然解约等。

而想要从分析现状到精准预测未来，就必须将分析方法从点(交易数据)深化到线(互动数据)。例如，亚马逊网站通过网页的点击串流数据，追踪使用者从进入到离开该网站的动线与行为，就是顾客与企业网站之间的互动数据。如果从中发现多数使用者点入某个页面就跳开，则代表该页面需要改善，让使用者在浏览网页的过程中没有压力或挫折感，能以最少的力气发挥最大的效能。

(3) 传感数据。传感数据即为GPS、RFID、视频监控等物联网设备带来的传感数据。在微处理器和传感器变得越来越便宜的今天，全自动或半自动(通过人工指令进行高层次操作，自动处理低层次操作)系统可以包含更多智能性功能，能从其环境中获得更多的数据。随着现在系统设计所包含的传感器和处理器越来越多，传感器和处理器价格的不断降低，人们在越来越多的系统或场合中将会自动地产生传感数据。

4. 大数据的产生原因

人类历史上从未有哪个时代像今天一样产生如此海量的数据。数据的产生已经完全不受时间、地点的限制，数据的总量在不断地增加，增加的速度也在不断地加快。大数据的成因，不仅是人类信息技术的进步，而且是信息技术领域不同时期多个进步交互作用的结果。大数据产生的原因主要来自以下4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、存储硬件的体积日益减小。1965年，英特尔(Intel)创始人之一戈登·摩尔(Gordon Moore)提出著名的摩尔定律，即：当价