

# 大数据基础知识

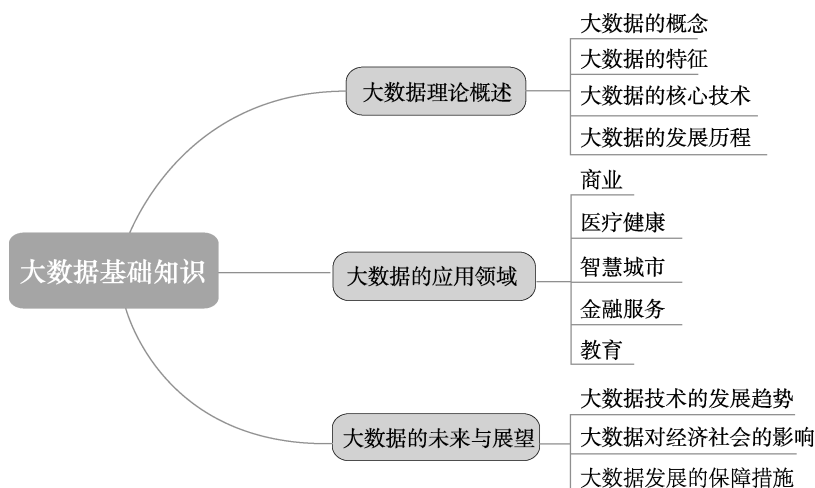
## 本章导读：

随着信息技术的高速发展，数据已成为经济社会中不可或缺的重要资源。大数据作为信息技术领域的重要分支，正在从多个方面改变着人们的生活和工作方式，并推动着各行各业的变革与创新。本章旨在通过介绍大数据的基本概念、特征、核心技术、发展历程及应用领域等，引导学生深入理解大数据的相关基础知识，帮助学生构建大数据思维，为学生的未来职业规划打下坚实的基础。

## 本章学习目标：

1. 掌握大数据的基本概念和特征。
2. 掌握大数据的核心技术内涵。
3. 熟悉大数据的应用领域与实践案例。
4. 了解大数据技术的最新发展趋势。

## 知识框架图：



## 1.1 大数据理论概述

### 1.1.1 大数据的概念

不同的研究机构对大数据的概念界定有所不同。

高德纳咨询公司认为，大数据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的海量、高增长率和多样化的信息资产。

麦肯锡全球研究所指出，大数据是一种规模大到在获取、存储、管理、分析方面超出了传统数据库软件工具能力范围的数据集合。这个定义主要突出大数据在规模上的巨大挑战，以及传统数据库软件工具在处理大数据时的局限性。

本教材认为，大数据是指无法在一定时间范围内用常规软件工具进行采集和处理的数据集合，是一种具有一定规模的、有分析价值的、能够反映客观事实的信息资产。

### 1.1.2 大数据的特征

如图 1-1 所示，大数据有规模性（Volume）、高速性（Velocity）、多样性（Variety）、价值性（Value）和真实性（Veracity）5 个特征，具体如下。

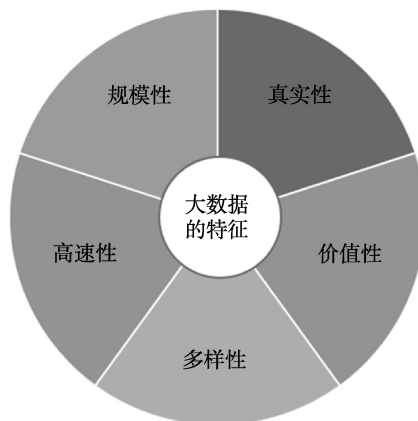


图 1-1 大数据的特征

#### 1. 规模性

大数据包含了大规模、海量的数据，这种数据量远远超出了传统数据库或常规数据处理工具所能处理的范围。大规模的数据量要求使用专门的存储和处理技术，如分布式存储系统和并行处理框架。

大数据的规模性主要体现在其数据量巨大。大数据包含了如今几乎所有领域的的数据，无论是社交媒体产生的用户数据、商业交易数据、互联网搜索记录，还是物联网、科学研究等产生的数据。

每个不同的领域都会同时产生大量的数据。例如，在社交媒体领域，如微信、钉钉、微博等社交媒体平台，每天都会产生数十亿条的用户交互数据，包括数以万计的帖子、图片、视频、点击量和浏览记录等。这些数据量十分庞大，并呈现持续增长的态势，是大数据的重要组成部分。在商业领域，如淘宝、抖音商城、京东、亚马逊等电商平台，每天都会产生大量的交易，这些交易数据都会被记录下来。这些交易数据数量巨大，对于电商企业来说十分重要，能够为日后客户偏好分析提供数据支撑。在金融市场领域，如股票、债券、期权、期货市场的交易也会产生大量的市场数据和交易数据，这些交易数据能够为市场参与者作出交易决策提供参考。

## 知识链接1-1

### 何为大数据之“大”？

在计算机科学中，数据的大小通常以存储单位来衡量。数据在计算机系统中采用二进制表示，1位二进制数“0”或“1”存储称为1 bit（位）。Byte（字节）是数据存储的基础单位。以下是常见的存储单位及其换算关系。

8 bit = 1 Byte（一字节）。

1 024 B = 1 KB（KiloByte，千字节）。

1 024 KB = 1 MB（MegaByte，兆字节）。

1 024 MB = 1 GB（GigaByte，吉字节）。

1 024 GB = 1 TB（TeraByte，太字节）。

1 024 TB = 1 PB（PetaByte，拍字节）。

1 024 PB = 1 EB（ExaByte，艾字节）。

1 024 EB = 1 ZB（ZetaByte，泽字节）。

1 024 ZB = 1 YB（YottaByte，尧字节）。

1 024 YB = 1 BB（Brontobyte，珀字节）。

1 024 BB = 1 NB（NonaByte，诺字节）。

1 024 NB = 1 DB（DoggaByte，刀字节）。

为了形象理解这些单位的大小，可以将其与现实生活中的例子进行对比。

1 Byte：1个英文字母。

1 KB：大约相当于四五百个汉字的短篇小说。

1 MB：大约相当于一部10 min的高清视频通话所消耗的数据流量，或者一张高清手机壁纸的大小。

1 GB：大约相当于一部2 h的高清电影的存储容量。

1 TB：大约相当于一个家庭一年内拍摄的所有照片和视频的总容量，或者约1 000部高清电影的存储容量。

1 PB：大约相当于全中国学术研究图书馆藏书信息内容的50%。

1 EB：大约相当于至今全世界人类所讲过的话语的20%。

1 ZB：大约相当于全世界海滩上沙子的数量总和。若在线数据以50 MB/s的速度进

行下载，大约需要 71.4 万年。

1YB：大约相当于 7 000 位人类体内的微细胞总和。这个容量已经远远超出了目前人类社会的数据存储需求。

1BB：大约可容纳 10 万亿人的基因组信息（远超地球总人口）。

1NB：大约相当于以当前光纤极限速率 100 倍的速度，光速传输超过 1 000 年的数据量，或质量接近小型天体的电子存储介质容量。

BB 和 NB 量级属于超大规模数据单位，现实中极少直接应用，更多用于科学假设或未来技术展望（如量子计算、星际数据网络）。

从个人日常的数字消费到全球信息的海量积累，大数据的规模在不断扩展，其在各个领域的应用也日益广泛，深刻地影响着我们的生活和工作。

资料来源：根据网络公开资料整理而成。

## 2. 高速性

大数据的高速性体现在数据的生成、传输和处理速度极为迅猛。例如，在金融市场，股票交易数据的实时更新速度可达每秒数百万笔；在社交媒体平台，用户生成的内容和互动数据也在分秒间海量累积。这种高速的数据流动，对数据的实时收集、传输和存储提出了极高的要求，传统的数据处理技术难以应对如此快速的数据流。因此，大数据技术必须具备强大的实时处理能力，能够迅速对海量数据进行清洗、分析和挖掘，从而及时提取出有价值的信息，为决策者提供实时、准确的洞察。例如，在自动驾驶领域，车辆通过传感器实时收集道路、车辆和行人数据，大数据技术的高速处理能力确保了车辆能够快速作出反应，保障行驶安全。高速性是大数据区别于传统数据的重要特征之一，它使得大数据处理能够捕捉到瞬息万变的信息动态，为各行业在快节奏的环境中把握机遇、应对挑战提供了有力支持。

## 3. 多样性

大数据有多种数据类型，如文本、图片、音频、视频等。常见的文本格式有 TXT、PDF、DOCX 等；图片格式有 GIF、JPG、PNG 等；音频格式有 WAV、MP3、CD 等；视频格式有 AVI、MP4、MPEG 等。大数据还有多种数据形态，其中包括传统的结构化数据，如数据库中的表格；半结构化数据，如 XML 文件；非结构化数据，如社交媒体帖子、图像和视频等。非结构化数据要比结构化数据更加难以储存和处理。因此，大数据的多样性要求数据处理和分析技术能够处理不同类型、不同形态的数据，并从中提取有价值的信息加以利用。

例如，阿里巴巴不仅拥有来自淘宝、天猫等自有电商平台的用户行为数据和交易数据，还通过支付宝等金融服务获取用户的支付数据，此外，阿里巴巴还通过与社交媒体、物流公司等合作伙伴的合作，获取更广泛的数据资源。这些数据中不仅包含了结构化数据（如用户注册信息、订单信息等），还包括半结构化数据（如商品描述信息等）和非结构化数据（如用户评价、社交媒体讨论等）。

#### 4. 价值性

大数据的价值性主要指大数据在各个领域所能产生的实际效益和潜在影响。尽管大数据中包含大量的信息，但并不是所有数据都具有同样的价值。数据处理和分析可以从大数据中提取出具有业务价值的信息，帮助企业作出更明智的决策、优化业务流程、提高市场竞争力、创造新的商业模式等。

例如，亚马逊通过收集和分析用户的购买行为以及与网站互动等数据，能够实现数据驱动的决策和流程优化。其推荐系统、需求预测、A/B 测试和移动 App 数据收集等都是数据驱动战略的体现。亚马逊的推荐系统能够基于用户的购买历史和浏览行为，为用户推荐可能感兴趣的商品，从而提高销售额和用户满意度，这体现了大数据在电商领域的应用价值。

#### 5. 真实性

大数据的真实性主要强调数据的准确与可靠。真实的数据必须准确反映客观世界中的事件和行为，与实际情况相符，没有经过人为篡改或歪曲。同时，真实的数据不仅要准确，还要具有可依赖性。这意味着数据要能够在多次验证和检验中保持一致性和稳定性，能够经得起时间和实践的考验。

例如，京东在智能库存管理中，通过区块链技术对交易数据进行上链存证，确保每一笔订单数据不可篡改；同时，对用户浏览行为数据采用多重加密校验，防止数据在采集、传输过程中被恶意篡改。此外，系统会定期将实时采集的数据与历史基准数据进行比对，若发现异常波动（如某类商品浏览量突然激增），会触发人工复核机制，确保数据的准确性与可靠性。这种对数据真实性的严格把控，为库存预测提供了可靠的基础。

### 1.1.3 大数据的核心技术

如图 1-2 所示，大数据的核心技术包括数据采集与清洗、数据存储管理、数据挖掘和数据可视化等方面，为各行各业提供了强大的数据支持和分析能力。



图 1-2 大数据的核心技术

#### 1. 数据采集与清洗

数据采集是指通过特定手段和方法，从各种数据源中收集并获取所需数据的过程。常见的数据采集工具包括数据采集器（如传感器、摄像头等）、日志收集系统（如 Flume、Scribe、Rsyslog 等）和各种网络爬虫。

数据清洗是数据处理与分析过程中的一个关键步骤，它涉及从原始数据中移除错误、重复或不完整的数据等，以确保数据的准确性、一致性和完整性。一般的数据清洗主要针对残缺数据、错误数据和重复数据进行处理。

## 2. 数据存储管理

数据存储管理是指将数据以某种格式保存在物理或虚拟的存储介质中,以便在日后需要时能够进行访问、检索和使用。数据存储管理对于确保数据的可用性、持久性和安全性至关重要。大数据是实时产生的,并且产生速度很快,比如非结构化数据通常以每年增长 50%的速度快速激增。随着科技的进步,越来越多的传感器采集数据,各种移动设备和社交多媒体也在产生数据。如此海量的数据需要高性能、高吞吐率、大容量的基础设备来存储。常见的存储方式有本地存储、云存储、数据库存储、分布式存储、内存存储和外部介质存储等。

## 3. 数据挖掘

数据挖掘(Data Mining)的核心目标是帮助企业 and 组织从海量数据中获取有价值的信息,以支持决策制定和预测未来趋势。数据挖掘与分析经常使用的工具有 SPSS 及 SAS、KNIME、RapidMiner、Weka 等,经常使用的编程语言有 Python、R。

## 4. 数据可视化

数据可视化是指通过可视的、交互的方式展示数据的一种方法。数据可视化不仅包括图表、图形、图像呈现,任何能够借助于图形的方式展示事物原理、规律、逻辑的方法都属于数据可视化的范畴。利用直观可视的方式呈现数据,观察者能够更加简单明了地接收和理解信息。通过图形化手段可以将数据中不可见或难以分析的信息转化为可观测的图形或符号等形式,从而提升识别和传递有用信息的效率。

常见的数据可视化呈现方式有折线图、散点图、关系图、热力图、地理空间图、词云图、三维图、网络图等。为了呈现更好的效果,可以根据具体需求和场景选择合适的工具。数据可视化工具的类型主要包括传统可视化工具(如 Excel)、商业智能(Business Intelligence, BI)工具(如 Tableau、Power BI)、基于编程语言的数据可视化(如 Python、R、JavaScript),以及在线数据可视化(如 Datawrapper、Google Charts 等)。

## 知识链接1-2

### 阿里云“云原生一体化数仓”

随着企业商业模式的数字化转型,企业对大数据技术的需求也日益增长,传统的数据库已经无法满足企业的需求。阿里云为了迎合市场需求,推出了“云原生一体化数仓”产品,为客户企业提供一站式、全链条的大数据平台。

阿里云“云原生一体化数仓”将 MaxCompute、DataWorks、Hologres 三种产品集于一体,实现了离线实时一体化、湖仓一体化和分析服务一体化的数据处理方式。这种多方位一体化的处理方式打破了传统数据库的局限性,为企业客户提供了更加灵活和多样化的数据处理能力。“云原生一体化数仓”能够离线处理和实时分析相互切换,满足企业客户不同情况下对数据的需求。同时,企业客户可以在同一平台上对结构化数据和非结构化数据进行处理,提高了数据的使用效率和分析能力。“云原生一体化数仓”还能够利

用建模能力为企业客户实现数据的全生命周期管理,以此来保证数据的一致性和完整性。

阿里云“云原生一体化数仓”现已被多个行业使用,包括工业、电子商务、物流、金融等,体现了大数据核心技术在企业数字化转型过程中的重要作用。

资料来源:阿里云官方网站。

## 1.1.4 大数据的发展历程

### 1. 萌芽阶段 (1980—2008 年)

“大数据”一词首次被提出是在 1980 年,美国著名思想家、未来学家阿尔文·托夫勒在其著作《第三次浪潮》中将其称赞为“第三次浪潮的华彩乐章”。1997 年,美国国家航空航天局(National Aeronautics and Space Administration, NASA)艾姆斯研究中心的戴维·埃尔斯沃斯和 NVIDIA 的迈克尔·考克斯在针对数据可视化问题的研究中,首次使用“大数据”概念。他们认为,面对复杂且海量的数据,信息技术的发展一定会使数据处理技术得以应用。2008 年,随着互联网的兴起和数据的爆炸性增长,《自然》杂志推出了名为“大数据”的封面专栏,用于探讨大数据科学技术及其在多个领域的应用。这一专栏的推出,象征着“大数据”概念已经被经济社会普遍认同。随着数据挖掘理论和数据库技术的快速发展,一些商业智能工具和知识管理技术也逐渐开始被应用于社会经济发展中,如数据仓库(Data Warehouse)、专家系统、知识管理系统等。这些技术的创新发展为大数据时代随后的发展和繁荣奠定了基础。

### 2. 发展阶段 (2009—2012 年)

随着互联网技术的发展,网络数据呈现指数级增长,大数据技术逐渐被广泛使用。2009 年,大数据技术开始被政府部门所使用,如印度政府利用大数据技术建立生物识别数据库,美国政府通过数据网站的方式对政府数据进行了共享和使用。2010 年,肯尼斯·库克尔在《经济学人》杂志上发表了大数据专题报告《数据,无所不在的数据》,对大数据的潜力和挑战进行了深层次的讨论。Google、Amazon、Facebook(现已更名为 Meta)、Twitter(现已更名为 X)等多家大企业利用大数据技术研发出各种功能性工具和应用,推动了大数据产业生态系统的不断发展和壮大。

“大数据”成为热门词汇在各个领域被广泛熟知。2011 年,麦肯锡在对大数据的研究报告中正式对“大数据”的概念进行了界定,并指出“大数据时代已经到来”。麦肯锡全球研究院发布的《大数据前沿报告》,详细介绍了大数据在各个领域的应用以及大数据的技术性框架。国际数据公司(International Data Corporation, IDC)的研究结果表明,截至 2012 年,随着大数据市场持续扩张,全球数据量已达到惊人的规模,从 TB 级别跃升到 PB、EB 甚至 ZB 级别。舍恩伯格《大数据时代》的出版,在全球范围内引起强烈反响,极大地推动了“大数据”概念在世界范围的普及和应用。随着大数据的影响日益扩大,我国也开始关注大数据产业的发展。时任中国科学院院长白春礼院士,在多次论坛和研讨会中呼吁应将大数据上升为国家战略,为后续中国大数据相关政策的推动奠定

了基础。

### 3. 成熟阶段（2013—2015年）

2013年被誉为“世界大数据元年”，此时大数据相关技术已进入成熟阶段。国内企业和政府开始利用大数据在多个领域进行决策和应用。2013年9月，贵阳联合中关村共同打造了“中关村贵阳科技园”。证券市场监管部门也在2013年9月开始利用大数据技术对证券市场内幕交易、市场操纵等违法行为进行监控，并通报马某利用多个股票账户进行违规操作，累计交易额10.5亿余元。同年11月，国家统计局协同腾讯、阿里巴巴、百度等11家企业签署了战略合作框架协议，将大数据技术应用于政府统计中。

在我国2014年的《政府工作报告》中，大数据被纳入国家战略发展层面。2014年3月，第四届中国数据中心产业发展联盟大会暨IDC产品展示与资源洽谈交易大会召开，贵阳被评为“最适合投资数据中心的城市”。同年8月21日，“2014贵阳云计算——大数据高峰论坛暨大数据产业技术联盟揭牌仪式”在贵阳隆重召开。同年10月20日，国内首款大数据指数基金开售，引起了各界投资人士的广泛关注。

2015年4月14日，全国首个大数据交易所——贵阳大数据交易所正式挂牌运营，并完成了首批大数据交易，这标志着大数据开始在金融领域发挥重要作用。同年9月，国务院发布《促进大数据发展行动纲要》，积极全面推进我国大数据发展和应用，进一步提升创新创业活力和社会治理水平。同时，在“十三五”规划建议中明确提出要实施国家大数据战略。

## 知识链接1-3

### 数字经济与国家战略——从“东数西算”看贵州发展新篇章

自2022年2月我国“东数西算”工程启动以来，贵州作为八大国家算力枢纽节点之一，其贵安数据中心集群已成为十大国家数据中心集群之一，这为贵安数字经济的发展带来了重大机遇。贵安新区作为全国首个大数据综合试验区的核心区，已经取得了显著成就。

据2024年中国国际大数据产业博览会消息，贵安在数字经济增加值方面取得了新突破，2023年数字经济增加值占国内生产总值（Gross Domestic Product, GDP）比重已提升至52.9%，电子信息制造业总产值同比增长29.3%，软件和信息技术服务业收入突破800亿元，同比增长20.6%。这些成果标志着贵安在“一硬一软”两大产业的发展上取得了显著成效。

贵安数据中心集群作为国家枢纽节点十大数据中心集群之一，对标全国一体化大数据中心，协同创新体系要求，致力于打造面向全国的算力保障基地。2023年，新建成投运了网易、美的、国电投3个数据中心，同时还有9个数据中心在建及签约待建，建成标准机架13.7万架，服务器承载规模达137万台，上电服务器81.1万台。这些数据中心的建设和运营，为“东数西算”工程提供了坚实的基础。

此外，贵安新区还加快了数字产业和人才聚集区、数字场景应用示范区、生态文明

建设展示区的建设，设立了 150 亿元的贵阳大数据科创城产业投资基金，进一步推动了数字经济的发展。

“东数西算”工程的稳步推进，不仅提升了贵州在全国的算力地位，也为贵安的数字经济发展注入了新的动力，推动了地区经济的高质量发展。随着数据中心集群的不断扩大和算力的增强，贵安有望成为全球超大型数据中心集聚度最高的地区之一，这进一步巩固其在数字经济领域的地位。

资料来源：中国政府网。

#### 4. 大规模应用阶段（2016 年至今）

自 2016 年至今，华为、阿里巴巴、腾讯等知名企业在大数据技术研发和应用方面取得了显著成果，并引领着大数据行业的发展方向。大数据与 5G、云计算、人工智能（Artificial Intelligence, AI）等新一代信息技术相融合，推动了数字经济的蓬勃发展。大数据的应用已经渗透到社会中的各行各业，如政府服务与公共治理、商业、金融、交通、通信、空间地理等领域，推动了这些行业的数字化与智能化应用创新。

我国相关政府部门出台了一系列政策措施来支持大数据产业的创新和发展。例如，国家发展改革委、工业和信息化部、商务部等部门联合主办的中国国际大数据产业博览会，推动了大数据技术的交流和合作。在“十三五”和“十四五”发展规划中，大数据均被列为重要的发展方向，大数据标准体系的完善成为发展重点。

经过多年的发展，大数据产业规模持续快速增长。根据中国信息通信研究院发布的统计数据，2022 年大数据产业规模达 1.6 万亿元左右，同比增长 20.8%；2023 年大数据产业规模在 1.9 万亿元左右，同比增长达 18.8%。北京、上海、贵州等地区凭借中央政府和地方政府提供的政策优势、资源倾斜以及产业基础设施建设，已成为大数据产业发展的重要地区，在大数据产业发展方面取得了显著成果。

## 1.2 大数据的应用领域

### 1.2.1 商业

#### 1. 市场营销

相较于传统的营销模式，大数据可以使市场营销更具针对性。

大数据通过收集和分析消费者的购买信息，如购买记录、浏览历史、地理位置等数据，能够帮助企业绘制详细的用户画像。利用用户画像，企业能够更准确地识别目标客户群体，从而实现精准营销。例如，通过获取和分析消费者的购买记录和商品偏好，企业可以向其推送符合其偏好的产品。基于大数据的用户画像，企业还可以制定个性化的营销方案，为不同用户提供符合其特点的优惠活动和广告，提高用户的购买欲望。这种个性化的营销方式不仅提升了用户的满意度和用户黏性，还增强了营销效果，从而帮助企业提高收益。同时，大数据分析工具可以根据历史数据来预测营销活动的效果，帮助

企业提前制定和优化营销策略，或进行消费数据实时跟踪，使企业能够及时了解营销效果并调整营销策略，以提高营销效率，增加营销收入。

除此之外，大数据还可以帮助企业利用互联网快速收集和分析竞争对手的信息资源，包括产品定价、市场份额、销售情况等。这些信息能够帮助企业及时了解竞争对手，制定具有竞争力的营销策略，从而抢占市场，增加销售份额。

## 知识链接1-4

### 阿里妈妈发布 2024 “双十一” 白皮书——淘宝的用户画像

在“双十一”购物节来临之际，淘宝的用户画像技术发挥着至关重要的作用。该技术通过大数据分析消费者的特征和需求，为商家制定精准的营销策略提供有效依据。

淘宝的用户画像包括消费者年龄、性别、消费水平、兴趣爱好、购买习惯等多种信息。利用精细化、精准化的数据分析，商家可以更加准确地把握消费者的需求和市场趋势。通过人工智能算法对消费者的数据进行深度学习（Deep Learning）和分析，商家可以

以更加精准地预测消费者的购买行为和需求，来提供更加个性化的营销和服务。例如，对年轻消费者推出更加时尚、个性化的商品；对家庭主妇推出实用、性价比较高的家用商品等，并推送一些优惠信息，吸引客户消费。

根据淘宝的用户画像，商家可以将广告投放到最有可能购买的用户群体中，提高广告的投放效率和交易成功率。这种精准的广告投放能够有效节省广告费用，减少无效投放比率。

资料来源：新浪网。



扩展资源 1-1 阿里妈妈  
2024 “双十一” 白皮书网  
址

## 2. 产品研发

在产品研发方面，大数据利用其庞大的数据量和强大的分析能力，为企业的产品研发决策提供有力的数据支持。

通过对历史研发项目的相关数据进行深入分析和挖掘，企业可以评估不同研发项目的各方面进展情况，可以发现其研发过程中出现的问题，并针对这些问题作出有针对性的改进方案，可以大大降低研发成本、提高研发效率。利用大数据分析与挖掘，企业能够发现新的市场机会，并预测产品市场的未来发展趋势，为产品研发提供新的思路和发展方向。同时，企业还可以通过分析竞争对手的产品数据和市场表现，及时调整自己的研发方向和重点，从而确保自身产品的竞争优势。

此外，企业可以在产品上市前进行用户试用测试，收集用户反馈和产品性能的大数据，及时发现产品存在的问题并进行优化改进。在产品上市后，企业也可以通过大数据对产品性能进行实时监控，及时发现问题并提升产品功能，提高产品质量和用户满意度。

## 3. 供应链管理

在供应链管理方面，大数据能够利用历史数据帮助企业对生产计划、库存情况和采

购策略进行合理规划。同时，大数据技术还可以实现对供应链中的各个环节实行实时监控和数据分析，使企业能够全面了解其存货情况、订单配送状态、生产进度等信息。例如，通过对库存水平进行实时监控和预警，企业可以及时调整库存量，保证充足和合理的库存量，这样能够减少企业库存积压或缺货现象的发生，提升库存周转效率，有助于企业降低库存成本，提高库存的管理效率。

利用资源的共享和交流，大数据可以帮助供应链上下游企业之间加强合作，减少由于信息传递的延迟和误差引起的损失，提高整个供应链各企业间的响应速度和灵活性。

同时，利用地理信息系统和实时交通数据的结合，企业可以制定最优的运输路线和计划，降低运输成本，提高运输效率。通过在运输车辆、仓库货架等设备上安装传感器，企业可以实时接收物流数据并实时监控。如果发现异常情况，如运输延误、货物损坏等，系统可以及时发出预警，使企业迅速采取应对措施。

## 1.2.2 医疗健康

### 1. 临床诊断

在传统的疾病治疗过程中，医生的诊断决策主要依赖于自身临床经验、患者的病情自述以及各项医疗检查结果。然而，这种诊疗方式往往受医生经验的局限性、患者信息的不完整性以及疾病的复杂性等因素影响，使得诊断结果可能出现一定偏差。

为了提升诊断的准确性和治疗的有效性，大数据技术开始应用于医疗健康领域。通过深入分析患者的病历、医学影像、症状描述、药物过敏史、家族遗传病史及各项检查结果，大数据技术能够为医生提供全面且细致的患者信息支持，辅助医生作出更为准确的疾病诊断，并据此制定更为科学合理的治疗方案。

大数据不仅影响着疾病的诊断与治疗，还深刻改变了人类对健康的理解与管理方式。结合医疗记录、基因信息、生活习惯及饮食规律等多维度数据的深度分析，人们能够实现对疾病的早期预测与防控。通过对临床数据的挖掘，大数据技术可以揭示特定症状与疾病间的关联，评估不同治疗方案的效果差异，为医生提供更为可靠的决策依据。

此外，大数据与遗传信息的结合将为患者提供个性化的医疗方案。例如，在基因引起的癌症治疗中，根据肿瘤的遗传特征，大数据技术可以为患者量身定制有效的治疗方案。

大数据与人工智能算法的融合，进一步提升了诊断的精确度。大数据技术通过对医学影像（如 X 光片、CT 扫描、MRI 等）进行高精度分析，可以识别出微小的病灶，甚至发现肉眼难以察觉的早期病变，显著提高诊断的敏感性和特异性。在肺癌筛查中，人工智能辅助的 CT 影像分析技术能够有效识别早期肺癌症状，使患者有机会接受更早、更有效的治疗。

### 2. 个性化医疗

个性化医疗又称精准医疗，即利用患者病历、生理指标、基因信息等多维度数据信息，为不同患者量身定制更具有针对性的治疗方案，提高治疗效果，减少副作用。个性化医疗的核心在于“精准”。它不仅关注患者疾病的普遍特征，还关注患者的个体差异情

况。通过收集和分析患者的相关数据，主治医生能够更加详细、深入地了解病人的患病程度和特征，进而制定出符合患者自身条件的治疗方案。

在癌症治疗过程中，个性化医疗的应用尤为突出。比如，传统的癌症治疗过程，会对所有患者使用相同的化疗药物或放疗方案，但是由于病患个体存在差异，这会导致治疗效果不同，甚至可能引发严重的副作用。如果制定个性化的医疗方案，通过对患者的基因、病理特征进行深入分析，就能够精准地选择适合的靶向药物或免疫治疗方法。例如，有的患者属于基因突变引起的病症，对靶向药物会特别敏感，这种情况就可以使用靶向药的治疗方案；而有的患者不属于基因突变引起的疾病，则他们可能对靶向药物治疗无反应甚至产生副作用或耐药性，因而使用靶向药的治疗方案对这些患者并不适用。这种个性化的精准治疗方式，能够有效地提升治疗效果，减少不必要的药物副作用，从而提高患者的生活质量。

### 3. 药物研发

医学文献在药物研发中起到至关重要的作用，其中涵盖着过去学者的研究成果、实验方法以及经验教训等信息。大数据技术可以从海量的文献中快速寻找到所需的有效信息，为新药研发提供理论支撑和研究方向。相比于传统的文献分析法，这种数字化的分析方法提高了信息检索的效率和准确性，从而帮助科研人员快速定位研究前沿，提高劳动效率，使得药物研发进程加速。

除对医学理论进行研究外，也可以利用大数据技术有效分析临床药物的使用情况。临床数据能够直接反映出药物有效性和安全性。通过大数据分析，可以对临床试验数据进行深度挖掘，发现相关联的隐藏情况。而且不同的患者对药物的反应可能是不同的，临床数据能够详细记录不同患者群体对药物的反应差异、剂量-效应关系等。这不仅有助于优化临床试验的设计，还能较早地识别出潜在的风险，提高药品研发的成功率。同时，可以对药物投入市场后的患者使用情况进行分析，进一步评估药物在实际应用中的表现，为药物监管和临床决策提供科学依据。

在药物研发的过程中，药物的副作用不可避免。大数据技术能够更加及时和全面地检测药物对患者的使用效果和副作用。通过建立大规模的药物监测系统，收集患者用药后的反馈数据，结合患者身体健康情况记录等多种相关数据，可以对药物的疗效和安全性进行实时监测和深度分析。这有助于早期发现药物的潜在风险，及时调整治疗方案，同时也会为药物研发人员提供有效的反馈信息，促进药物的持续优化和改进。

分析大量医学文献和临床数据，有助于缩短新药研发时间、提高研发成功率、降低研发成本。有调查发现，使用大数据技术能将药物发现、临床前研究的时间缩短近 40%，为临床试验阶段节约 50%~60%的时间。利用大数据技术对药物的使用效果和副作用进行实时追踪和深度分析，医药企业能够根据相关数据为药物研发进行优化，为企业研发成果提供有力的数据支持。

### 4. 健康管理

随着科技的不断发展，智能可穿戴设备已经逐渐成为人们日常生活中重要的一部分，

如智能手表、健康手环、智能眼镜等。这些智能可穿戴设备能够实时追踪和记录用户的各种健康数据，如日常运动量、心率、血压、睡眠质量等，还能通过云端技术与大数据分析相结合，为用户提供个性化的健康管理方案。

智能可穿戴设备通过内置的传感器，能够全天候、不间断地监测用户的健康指标。例如，通过睡眠监测查看用户的睡眠质量和睡眠周期，通过分析深度睡眠、浅度睡眠或清醒时的状态，为用户提出具有针对性的提升睡眠质量的建议。除此之外，这些智能可穿戴设备还可以根据用户的运动情况，记录步数或消耗的卡路里数等，帮助用户了解自身的运动习惯并制订符合自身情况的运动计划。这些数据可以通过蓝牙或 Wi-Fi 上传至手机 App 或云端。经过数据分析后，用户可以直观地看到自己的健康状况报告，以及系统出具的健康改善建议。对于那些有慢性病的人群，如高血压患者、糖尿病患者等，这些智能可穿戴设备能够帮助他们对身体健康情况进行日常管理，使其自我管理变得更加有效。

智能可穿戴设备不仅可以针对个人的身体健康进行管理，还可以利用大数据技术，汇总大量的用户数据，对不同区域、性别和年龄的人群健康情况进行分析，寻找一些影响健康的关键因素。比如，通过对比不同年龄段人群的血压数据，可以发现血压随年龄变化的规律，以及可能影响血压升高的生活习惯或环境因素；对于老年人群体，可以重点关注骨质疏松、心血管疾病等常见病的预防和管理；而对于年轻人群体，则可以加强运动锻炼和心理健康的引导。这种深入的数据分析有助于对特定人群或公共卫生进行有效管理，并提出预防疾病发展的合理策略和决策。

## 知识链接1-5

### 华大基因推出肿瘤基因检测

美国知名的好莱坞女星安吉丽娜·朱莉，在 2013 年经过检测时发现自身携带致癌基因——BRCA1 基因。为防止罹患卵巢癌，她于 2015 年切除了卵巢和输卵管。随着生物技术和大数据技术的发展，个体基因检测治疗疾病已经成为现实。目前，国内外已经有多家基因检测机构，如华大基因、贝瑞和康、23 and Me 等。

华大基因一直致力于肿瘤基因组学研究，已经研究出 20 多种癌症类型。其肿瘤基因检测服务采用高通量测序手段对来自肿瘤病人的癌组织进行相关基因分析，可一次性检测多个与肿瘤相关的基因，能全面、系统、准确地解读肿瘤与基因的关系。该检测机构通过患者血清、口腔黏膜数据、基因测序、体检数据、电子病历、遗传记录、地理区域以及生活条件等数据信息，结合药物基因组相关信息，辅助临床医生为癌症患者选择合适的治疗药物，制定完善的治疗方案，从而满足患者个性化的治疗需求，延长患者的寿命。

华大基因的肿瘤基因检测服务已经在多地开展应用，并取得了显著成效。例如，在结直肠癌筛查项目中，华大基因与多个地方政府和卫生健康部门合作，通过引入先进的肿瘤检测技术（如粪便 DNA 甲基化检测），为当地居民提供了便捷的筛查服务。这些项目成功检出了多名癌症患者以及癌前病变患者，实现了早期诊断、早期治疗的防治目标。

资料来源：转化医学网。

### 1.2.3 智慧城市

随着物联网技术的快速发展，大数据结合物联网加快了智慧城市建设的进程。物联网作为连接物理与数字的桥梁，为智慧城市建设提供了丰富的数据源，而大数据技术则为处理和分析这些海量数据提供了有力支撑。

通过物联网技术，城市中的各类设施和设备（如传感器、摄像头、射频识别设备等）被连接起来，实时收集并传输数据，大数据技术则对这些数据进行收集、处理和分析，为城市管理者提供决策支持。智慧城市的构建为市民提供更加便捷、高效的服务和设施，如智能交通、智能医疗、智能家居等，有效提升城市居民的生活质量和幸福感。大数据和物联网技术的结合，可以帮助城市管理者实时监测城市运行状态，及时发现问题并解决，提高城市管理效率。通过对城市各项数据的收集和分析，可以更加精准地预测和满足市民的需求，优化城市资源的配置。例如，多个城市利用大数据技术分析交通拥堵原因和规律，制定限行、限购等措施，以有效缓解交通压力；通过物联网技术建立智能停车平台，实现停车位的实时查询、预订和支付，为驾驶者提供便利。

除此之外，在对空气质量、水质、噪声等环境影响因素进行实时监测时，通过大数据技术，可以及时发现环境污染问题，制定有效的治理措施，保护城市良好的生态环境。

在公共安全领域，通过对部署监控摄像头、烟雾探测器、入侵报警器等设备的检测，可以使政府部门实时掌握城市的安全状况，及时发现并处理安全隐患。同时，大数据分析还可以帮助预测犯罪趋势，为警方提供决策支持。总的来说，大数据技术能够有效提升公共服务水平，通过整合政府各部门的数据资源，为市民提供更加便捷、高效的政务服务。

#### 知识链接1-6

##### 人工智能“眼睛”助力电动车的治理——深圳龙华区

##### 人工智能视觉分析管理平台

近几年，电动自行车凭借灵活、便捷、低成本的优势，已经逐渐成为深圳市民的重要交通工具。但是在给予便利的同时，电动自行车也给城市交通规划、道路建设和城市治理带来了挑战，致使涉及电动自行车的交通事故频发。

据统计，在深圳龙华区，电动自行车保有量已超80万辆。为了有效管理电动自行车，保障市民的出行和居家安全，龙华区联合多部门，实现对电动自行车进行全链条、全环节、全方位的专项治理。

龙华区政府利用“数字治理”人工智能支撑平台，对辖区内62所幼儿园和中小学出入口进行覆盖，通过监控实时抓拍并自动识别未戴头盔人员，对学生家长群体头盔佩戴率进行智能评测。大浪街道利用治理平台检测到不佩戴头盔、走机动车道、加装雨棚等违法事件，并进行算法智能检测，然后将结果反馈给执法部门，辅助其开展整治工作。发现违法违规事件后，通过大模型的自动分拨技术实现自动分拨，并以小程序、短信、

电话等方式实现劝导、非现场执法等事件处置。最后，基于大模型及大数据技术实现事件的评价、整体态势分析，以及提出有效的改进意见。基于物联网、大数据技术、人工智能技术，依托监控资源，利用图像识别、视频与射频识别等技术，龙华区实现了电动自行车从销售、生产、登记、使用、回收多场景的各类事件发现。

整体上，龙华区实现了从事件发现、事件分拨、事件处置到事件评价的全流程闭环应用，电动自行车乱象治理效果显著。在半年的时间里，街道内骑行家长头盔佩戴率由原来的60%提升至90%以上，大大提升了骑行家长和骑行学生的安全。深圳云天励飞技术股份有限公司公共安全行业专家尹义表示，接下来还会将电动自行车治理与视觉多模态大模型进行结合，做到成本更低、场景更丰富、实战效果更好。

据了解，自2023年7月起，基于“数字治理”人工智能支撑平台的智能化支撑和高性能算力支撑，龙华区新部署上线了电动自行车治理算法超15种，算法准确率达到85%以上。算法上线以来，平台累计发现各类事件396万件，日均事件量超2万起。以“一车一档”制为例，龙华区基于视频智能分析，对骑行人员的骑行车牌和骑行人脸进行抓拍，并通过车牌和人脸构建骑行车辆和骑行人员档案。目前，系统已接入高铁站、地铁站、汽车站、医院等重点场所视频资源超300路，通过联动城管，实现重点场所周围超1000起电动自行车乱停放事件的处理，挖掘非法运营骑行人员超100人。

资料来源：深圳政府在线官网。

## 1.2.4 金融服务

目前，大数据在金融领域的应用已经涵盖了信贷风险评估、供应链金融、股市行情预测、股价预测、智能投顾、骗保识别、风险定价等多个方面。例如，商业银行通过对内外部数据资源整合，利用大数据技术对客户信贷风险进行评估，提高了贷款审批的准确性和效率；证券公司和保险公司也利用大数据技术进行市场预测、投资策略制定以及骗保行为的识别等。

通过大数据技术对大量金融数据的分析，金融机构能够更科学有效地配置资源，提高运营效率和盈利能力。利用大数据技术，金融机构能够为客户提供更加个性化、精准的金融服务，并提高相关业务办理速度和处理效率。

同时，金融数据的敏感性要求金融机构必须保障数据隐私和安全，防止数据泄露和滥用。大数据技术能够帮助金融机构更准确地识别、评估和管理金融风险，从而提升金融机构管理的安全性和高效性。随着人工智能、区块链等技术的不断发展，大数据技术与这些新兴技术的融合也大大提升了金融机构抵御风险的能力。

在信贷风险管理方面，利用多种传感器或者多种渠道采集数据，信贷部门能够更加全面、准确、及时地掌握借款人的信息，降低贷款风险。同时，通过整合并分析申请人的信用历史、交易行为、社交媒体活动等多维度信息，可以构建更为精细的风险评估模型。这不仅提高了风险识别的准确性，还缩短了贷款审批时间，使得客户能够享受到更为便捷的信贷服务。

对于小微企业而言，大数据技术通过分析其经营数据、供应链关系等，为供应链金融提供了新的视角，解决了传统金融模式下因信息不对称而导致的融资难问题。

在资本市场，大数据技术正深刻改变着股市行情预测和股价预测的方式。通过对历史交易数据、宏观经济指标、新闻报道、社交媒体情绪等多源信息的综合分析，机器学习（Machine Learning, ML）算法能够捕捉到市场微小的波动趋势，为投资者提供更加精准的决策支持。在大数据技术的支撑下，证券公司能够快速收集和传输大量的信息，并设计出符合客户需求的产品组合。大数据加持下的智能投顾服务能够根据投资者的风险偏好、财务状况和投资目标，自动为其定制投资组合，实现资产的优化配置和动态调整，降低了投资门槛，促进了理财的差异化竞争。

## 知识链接1-7

### 大数据助力金融服务

一直以来，中小企业长期面临融资难、融资贵的问题，传统金融机构对中小企业信用评价难，贷款发放谨慎。阿里小贷利用阿里巴巴 B2B、淘宝、天猫、支付宝等电子商务平台的数据优势，瞄准小额贷款市场，为中小企业提供融资服务。阿里小贷积累了大量的信用数据及行为数据，引入网络数据模型和在线资信调查模式，通过交叉检验技术辅以第三方验证确认客户信息的真实性。这些大量数据为阿里小贷提供了丰富的信息资源，使其能够更准确地评估客户信用状况和贷款风险。

通过大数据技术，阿里小贷能够更精准地进行风险评估，通过分析用户在互联网平台上的浏览记录、消费行为和社交网络等数据，为用户绘制出更全面的画像，并评估其信用风险。这种多维度的信用评估方法使得阿里小贷能够更准确地判断用户的还款能力和风险等级，利用大数据挖掘和分析技术，准确把握小微企业的资信情况和信用等级，进而降低不良贷款的风险，并提高贷款的安全性。

同时，通过对大数据的实时监测和分析，阿里小贷可以及时了解用户的需求和行为变化，从而有针对性地调整借贷产品和优化业务流程。通过实时监控小微企业的经营轨迹和资金使用情况，阿里小贷能够及时发现潜在风险并发出预警。这不仅提升了用户体验和满意度，还通过数据的持续迭代和优化，不断完善其业务流程和风控模型。除此之外，当通过数据分析发现异常交易或还款行为时，阿里小贷可以及时采取相应措施，保障贷款资金的安全。

资料来源：百度百科。

## 1.2.5 教育

通过对学生学习背景和过程相关的各种数据测量、收集和分析，从相关数据中归纳分析不同学生的学习特点和学习情况，学校可以为学生提供个性化的学习支持。通过分析学生的学习数据和兴趣偏好，学校可以更精准地进行课程设计和资源分配，提高教学质量和效率。

在传统的教育模式中，教师常常使用“一刀切”的教学内容和方法，往往难以精准掌握每位学生的学习特点和进度。然而，有了大数据技术的帮助，利用线上学习平台，学校和教师可以快速了解学生的学习习惯、兴趣偏好、知识掌握程度等多维度信息。例如，结合学生在在线学习平台上的行为数据，如浏览记录、作业提交情况、视频观看时长等，平台能够自动识别出学生的学习特征。根据这些信息，教师可以为每位学生量身定制学习计划，推荐适合这些学生的学习资源和路径。这种个性化的学习支持不仅能够提高学生的学习效率，还能激发学生的学习兴趣和乐趣，帮助他们在自己擅长和感兴趣的领域深入探索，实现自我能力的提升。

在保障基本教育规模的情况下，大数据可以帮助高校实现差异化教学。一方面，教师可以因材施教，根据学生的不同需求推荐合适的学习资源；另一方面，可以形成大规模的在线教育模式，如 MOOC（慕课）平台突破了传统教育中实体教室的限制，能同时满足数十万学习者的学习需求，促进教育资源的优化配置和教育公平的实现。通过实时监测学生的学习表现和反馈，能够帮助教师及时发现学习者的学习问题、给予他们指导和支持，提高他们的学习动力和学习效果。

通过采集分析管理者、家长、教师、学生等多方面的行为数据，教育管理部门能够更加全面、准确地制定管理方案和决策，提高教育管理的科学性和有效性，从而全面提升教学管理质量，为学生、教师、家长等提供更好的支持。

## 知识链接1-8

### 青岛九中构建智慧大数据平台

青岛九中积极响应教育现代化的号召，成功建立起一套智慧大数据平台，如图 1-3 所示。该平台通过数据集成和赋能教育治理，全面提升学校的数字化赋能能力，为教育治理和教学改革注入了新的活力。

青岛九中通过智慧大数据平台，利用高清摄像头和声音采集麦克风等设备记录每堂



图 1-3 青岛九中 App 界面

课上教师的教学行为，在一段时间后对积累的数据进行分析总结和反馈，帮助教师识别哪些教学方法具有积极的促进作用。这种基于大数据的课堂分析技术，不仅提升了教师的教学效果，还促进了学生的个性化学习。教师可以利用大数据技术进行个性化作业设计，对作业难度、阅读材料分配、学科内容难度调节等进行自动分析和调整，从而帮助学生高效完成相应课程目标。同时，学校还依托大数据精准教学系统在学校开展各类考试，并生成考后报告，为学生提供个性化的学习建议。这种精准的教学模式有效提高了学生的学习效率和学习成绩。

智慧大数据平台帮助青岛九中打造全方位智慧校园新生态，学校整体实现了数字化管理，学生“露一次面”就可以进行多种功能的集成管理，如食堂打卡、宿舍考勤、请假流程等。此外，学校还更新了防火墙和上网行为管理系统，大范围优化无线网络覆盖，为日常教学提供更加稳定的支持。这种智慧校园生态的构建，不仅提升了学校的管理效率，还使得学校的日常运作更加智能化和无缝化。

青岛九中被评为智慧校园“引领型”学校，并在山东省第二批教育信息化试点验收中获评“教育信息化示范单位”。这些荣誉和成果充分证明了青岛九中在大数据应用和教育信息化方面的领先地位。

资料来源：青岛九中官方网站。

## 1.3 大数据的未来与展望

### 1.3.1 大数据技术的发展趋势

#### 1. 数据量的爆发式增长

随着互联网的普及和物联网的发展，数据量正在以惊人的速度呈爆发式增长。近年来，我国大数据市场规模持续增长。有关数据显示，我国大数据企业注册量由 2018 年的 2.95 万家迅速增长至 2022 年的 14.41 万家，年均复合增长率达 48.67%。2023 年，中国大数据市场规模达到 6 482.2 亿元，同比增长 15.1%。

根据《全国数据资源调查报告（2023 年）》的统计数据显示，2023 年全国数据生产总量达 32.85 ZB，同比增长 22.44%。这一数据持续增长的趋势主要得益于互联网、物联网、大数据等技术的迅速发展和普及，这些技术推动了国内各行各业数据的快速生成和传输。据 IDC 等权威机构预测，全球数据量将继续以惊人的速度增长，其中，非结构化数据占比将显著增加，到 2025 年，全球数据量将达到 175 ZB。

随着数据量的爆发式增长，如何高效、安全地存储和处理这些数据将成为巨大挑战。各国政府和企业需要投入更多资源用于数据存储与处理技术的研发，提升数据存储的容量、性能和可靠性，降低数据处理的成本和时间。同时，数据量的增长也带来了数据安全与隐私保护的问题，为了保障数据的安全和隐私，防止数据泄露和滥用，应积极推动建立健全的数据安全与隐私保护法规和标准体系，加强数据加密、访问控制等技术手段的应用，确保数据隐私和数据安全。

## 2. 人工智能与大数据的深度融合

人工智能的发展离不开大数据的支持。大数据为人工智能提供了丰富的训练数据，使其能够通过不断的训练进行自身性能的提升和优化。同时，人工智能技术的应用也使得大数据技术处理过程更加智能化和自动化。人工智能与大数据的深度融合能够有效推动社会经济的全面数字化转型，为经济社会发展创造更多的价值。

在技术层面，随着人工智能算法的不断进步，特别是深度学习、强化学习（Reinforcement Learning）等先进算法的应用，人工智能系统能够更好地处理和分析大数据。同时，针对一些特定领域和任务定制模型将进一步提高数据分析的准确性和高效性。随着实时数据的生成速度越来越快，人工智能与大数据的深度融合更加注重提升实时数据处理能力。

随着人工智能与大数据的深度融合，数据安全与隐私保护将变得越来越重要。通过更加严格的数据加密、访问控制等技术手段，使用户的数据安全和隐私得以保护。

## 3. 边缘计算的兴起

边缘计算是一种分布式计算模型，是将数据处理和计算资源放置在接近数据产生源头的边缘设备、传感器或用户设备上，以提供更快速、更具有时效性的计算和数据分析能力。这种计算模型能够有效减少数据传输延迟、提高数据安全性、减轻网络带宽压力，并能够支持实时决策与自主操作。

在未来技术发展趋势中，数据量激增对实时数据的存储和处理需求日益增强。传统云计算模式数据处理和隐私保护方面存在局限性，边缘计算应运而生。边缘计算将成为未来大数据领域的重要发展方向。通过将数据处理推向数据产生的地方，如物联网设备附近，利用边缘计算技术可以显著减少延迟和带宽使用问题，有效提升数据处理的速度和效率。特别是在智能制造、自动驾驶、智慧城市、医疗健康等领域，它们对低延迟、高可靠性的数据处理有着迫切的需求，而边缘计算能够满足这些行业的基本需求。

各国政府也高度重视边缘计算的发展，并出台了一系列政策措施以加快边缘算力建设、推动“云边端”算力泛在分布与协同发展。我国在《“十四五”国家信息化规划》和《“十四五”数字经济发展规划》中将边缘计算列为发展重点，并强调加强针对特定场景的边缘计算能力，优化算力布局与智能调度。中国报告大厅网讯指出，中国边缘计算的市场规模从2018年的177亿元增加至2022年的542亿元，期间年均复合增长率为32.3%。根据中商产业研究院发布的《2024—2029年中国边缘计算深度分析及发展趋势预测报告》显示，2023年我国边缘计算市场规模达732亿元，同比增长35.09%。根据迪赛顾问预测数据，2026年我国边缘计算市场规模将超过1300亿元。欧盟在2021年的《2030数字指南针：欧洲数字十年之路》中规划指出，截至2030年要部署10000个高度安全的气候适应型边缘节点。在2023年，欧盟委员会批准了一项12亿欧元的国家援助计划，以支持欧洲先进云计算和边缘计算技术的研发并进行了首次工业上的部署。此外，由意大利公司Engineering牵头，约40家欧洲企业和组织组成的联盟，启动了DOME（Distributed

Open Marketplace Europe) 计划, 该计划要在欧盟实施云和边缘服务的开放分布式市场。

## 1.3.2 大数据对经济社会的影响

### 1. 推动数字经济增长

大数据已经成为推动各行各业发展的重要引擎。通过对海量数据的分析, 企业可以更加准确地把握市场需求, 优化产品设计, 提高生产效率, 从而推动数字经济的快速增长。

2023 年以来, 我国 5G、人工智能等技术创新持续取得突破, 数据要素市场加快建设, 数字经济产业体系不断完善, 数字经济全要素生产率巩固提升, 支撑了我国新质生产力的积累壮大。据《中国数字经济发展研究报告(2024 年)》披露, 2023 年, 我国数字经济规模达到 53.9 万亿元, 较上年增长 3.7 万亿元。数字经济占 GDP 的比重达到 42.8%, 较上年提升 1.3%, 对 GDP 增长的贡献率达到 66.45%, 第一、二、三产业数字经济渗透率分别为 10.78%、25.03% 和 45.63%, 第二产业数字经济渗透率的增幅首次超过第三产业。

数字经济全要素生产率的增长是新质生产力的关键标志, 对经济增长的贡献十分显著。2023 年 10 月 25 日, 国家数据局在北京正式挂牌成立。国家数据局呼应了数据驱动发展的全球治理态势, 标志着推动数据要素市场建设, 要积极促进数据要素价值的持续释放。

数字经济通过扩大数字投入, 推动社会经济发展形成质的跃升。数字经济通过发挥有效的市场作用, 吸引市场主体积极参与市场竞争, 推动经济发展释放活力。同时, 数字经济通过发挥数字投资与数字消费效率, 可以有效扩大市场需求, 推动经济发展形成量的扩张。

### 2. 创新生产方式

传统的生产方式注重规模化和集中化, 而大数据的出现改变了这种情况。通过大数据技术, 企业可以实现产品的定制化生产, 减少资源的浪费, 提高资源的利用效率。

大数据正在深刻地改变着生产方式。在大数据时代, 消费者与生产者可以直接通过网络进行对接, 实现以销定产, 从而减少库存的积压, 实现精确生产和个性化生产。这种生产模式的变革, 要求生产者更加准确地掌握市场需求情况, 减少浪费的发生, 提高生产效率。数字化技术的应用使得消费者的需求成为生产者生产的主要依据, 个性化的智能生产和定制生产成为发展趋势。利用大数据分析, 生产者可以定位消费者的个性化需求, 从而实现定制生产, 满足市场的多样化需求。利用多维度的市场信息反馈, 生产者能够及时了解用户体验、竞争对手情况及同类产品信息, 通过数据分析优化当前的研发、生产活动, 增强整个生产过程的开放性和灵活性。这种生产模式有助于生产者不断优化产品, 提高自身市场竞争力。

### 3. 改善社会治理

大数据为政府机构进行社会治理提供了新的思路。政府利用大数据技术, 通过了解

社会需求，更有效地提供公共服务。同时，大数据分析还可以帮助政府发现社会生活中存在的问题，并采取针对性的措施加以解决。

大数据技术可以推进社会治理角度转换，即从结果治理转为预防治理。利用智能化的社会治理措施，以预防治理为切入点，积极推动社会安全建设，提高社会治理的可预见性，使政府的决策科学有效，从而提升政府行政能力。大数据技术的应用使得社会治理能更精准地定位问题、寻找原因并采取措施。

## 知识链接1-9

### 鄂尔多斯市康巴什区“多多评·码上生活”平台

鄂尔多斯市康巴什区自主研发了“多多评·码上生活”国有数字化综合平台（简称“多多评”平台），并率先试点推行“码上生活、码上服务、码上治理”的“二维码+”智慧治理模式。“多多评”平台部分功能区如图 1-4 所示。

康巴什区为下属的 4 个街道、19 个社区近 6 万户居民和商户定制了专属的“一户一码”。此外，“户内码”开通网上政务服务。“户外码”“商户码”通过“我要点赞”“我要上报”“我要联系”三大功能板块，形成有效的互动沟通模式。依托“多多评”平台，康巴什区设置干部单位码，为所有干部制作可随身佩戴的“服务评议卡”，引导广大各级干部自觉接受群众评价、主动征求群众意见，评价结果最终会纳入年终考核。

2022 年，“多多评”平台累计收集居民群众“急难愁盼”问题 3 400 余件，解决各类矛盾纠纷 2 300 余起，处置率和实时反馈率均为 100%，事件处理效率由以往 1.5 起/h 提高到 5.3 起/h。这种高效的处理方式，大大提升了政府在群众中的公信力。截至 2024 年 6 月，“多多评”平台在康巴什区注册用户已超 35.2 万人，日均访问量达 19 500 人次，单日峰值访问量最高达 35 万人次，累计处理接诉即办、服务事项 12 633 件。

鄂尔多斯市康巴什区通过“多多评”平台，以数字化、智能化手段进行基层社会治理，提高了社会治理的科学性、精准性和有效性。



图 1-4 “多多评”平台部分功能

资料来源：网易。

### 1.3.3 大数据发展的保障措施

#### 1. 统筹推进基础设施建设

对已经存在的基础设施要进一步完善，全面提升国家信息化水平和促进数字经济发展。同时还要规划部署好高速光纤网络、无线宽带网的建设，确保其在信息技术领域保持领先。

在云计算中心、数据信息资源平台和物联网基础设施建设中，要推动这些设施的后续应用，推动跨行业数据资源的高效采集、集成和共享。通过构建云计算中心，提升强大的数据处理和存储能力，支持各类应用的高效运行。数据信息资源平台则可以使数据更便于集中管理和共享，促进数据价值的利用。物联网基础设施的建设能够推动万物互联的实现，为智慧城市治理、智能交通发展等提供有力的支持。

在高速光纤网络和无线网络方面，需要加大资金和相关技术投入，引进和升级现有的光纤传输和无线通信技术（如 5G 技术），提高网络传输速度，以满足用户日益增长的需求。要优化网络结构，增强网络的稳定性和可靠性，确保数据安全、高效地传输。特别是偏远地区和农村，更加需要加强政策引导和网络通信技术的创新，增加 5G 网络的覆盖广度和深度，让更多人享受到高速、稳定、快捷的通信网络服务。

同时，还要深入推进“三网融合”（即电信网、广播电视网、互联网的深度融合），增强市场竞争力，提升电信服务的性价比和用户满意度。

#### 2. 加大政策与法规支持力度

积极完善的法律法规能够有效推动大数据产业的健康发展。要对数据产权明确其归属，合理界定数据的所有权和使用权，避免因产权不明确而引发的纠纷。在大数据法律法规的制定过程中，应当对数据交易的流程和规则进行规范性建设，确保交易的公平性和透明度，防止不正当竞争和市场混乱。还要防止数据滥用和隐私泄露等问题的发生，保护个人和企业的合法权益。在大数据的使用中，数据共享十分重要，要建立合理的数据共享规则，促进数据资源的充分利用和创新应用，推动各行各业的发展。

在政策支持与引导方面，政府可以通过出台一系列政策措施鼓励和支持大数据技术的发展和运用，比如提供财政补贴，以减轻大数据企业在研发和运营过程中的经济压力，降低企业的创新成本。政府还可以实施税收优惠政策，降低大数据企业的税收负担，从而激发市场活力，吸引更多企业投入大数据领域。

### 本章小结

1. 大数据是指无法在一定时间范围内用常规软件工具进行采集和处理的数据集合，是一种具有一定规模的、有分析价值的、能够反映客观事实的信息资产。
2. 大数据的特征为规模性、高速性、多样性、价值性和真实性。
3. 大数据的核心技术包括数据采集与清洗、数据存储管理、数据挖掘和数据可视化。
4. 大数据的应用遍布各行各业，包括商业、医疗健康、智慧城市、金融服务以及教

育等领域。在这些领域中，大数据不仅优化了决策过程，提高了服务质量和创新效率，还催生了新的商业模式。大数据已成为推动行业进步和社会发展的重要力量。

5. 大数据的未来发展趋势，包括数据量的爆发式增长、人工智能与大数据的深度融合、边缘计算的兴起等。这些技术发展趋势将进一步提升大数据的处理能力、应用范围和安全性，为大数据的未来发展注入新的活力。同时，也要关注大数据对社会经济的影响，明白数字经济发展的重要性。

### 思维拓展

1. 大数据在个人生活中的应用越来越广泛，请结合自己的日常生活，列举出至少三种大数据技术的应用场景，并分析这些应用给你带来了哪些便利或影响？

2. 随着大数据技术的发展，未来对相关专业人才的需求会有哪些变化？作为学生，应如何提前准备？

3. 大数据在社会治理中的应用前景如何？请结合具体案例，谈谈大数据如何帮助政府提高公共服务效率和质量。

4. 在大数据时代，传统企业应如何利用大数据技术进行转型升级？

### 实践作业

2024 中国国际大数据产业博览会（简称数博会）于 8 月 28 日至 30 日在贵阳成功举办，本次数博会以“数智共生：开创数字经济高质量发展新未来”为主题，聚焦行业最新发展趋势和前沿技术动态，举办专业展览，配套举办了成果发布、行业交流活动等一系列特色活动。

在专业展览版块，整体布展面积达 6 万平方米，围绕国家数据工作的开展重点，规划数字产业化、产业数字化、数据价值化、数字化治理、数字新基建、数据安全 6 大主题版块，同时设置了一批相关主题展区，包括算力生态、数据要素×智能制造、数据要素×数字化转型、人工智能等，共吸引国内外 414 家企业积极参展，在展区各参展企业将大数据领域的新技术、新产品、新方案、新应用 1 300 余项进行展出，推动了企业间的交流与进步。除此之外，在展会期间还举办了参展企业自主发布等活动，为企业提供了专业的展示平台。

8 月 29 日，国家数据局联合科技部、农业农村部、文化和旅游部、中国科学院、中国工程院、国家文物局、国家中医药管理局等部门在中国国际大数据产业博览会上发布第二批 28 个“数据要素×”典型案例。这批案例的发布，对解决数据行业发展难题和提升行业效益有重大作用，彰显了数据要素推动经济发展的乘数效应。

随着信息技术和数字经济的快速发展，大数据已成为社会发展创新的重要推动力量。在本次实践作业中，请同学们结合上述材料和其他相关知识，以小组的形式探讨大数据在你的生活中产生了怎样的影响。

**即测即练**

自  
学  
自  
测



扫  
描  
此  
码