



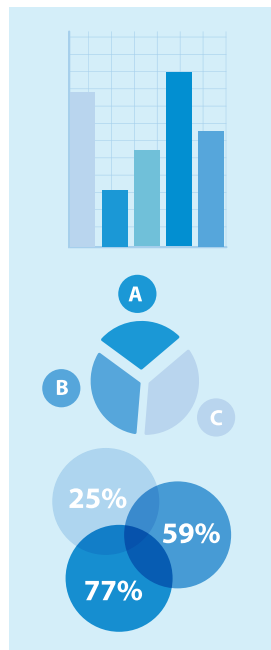
大数据 应用基础教程

刘三满 —— 主编

大数据
互联网+

价值引领
实践导向

校企双元开发
项目任务教学





本社微信公众号
(请用微信“扫一扫”)



本社天猫旗舰店
(请用手机天猫APP“扫一扫”)



定价：49.00元



“十四五”大数据融媒体新形态教材

大数据 应用基础教程

主 编 刘三满

副主编 刘荷花 王晓燕 陈 娜 程金凤 刘 珂 郭丽蓉

李 枫 朱飏凯 任 艳 郝红红

编 者 凌 晨 代依陈 李 馨



中国财经出版传媒集团
中国财政经济出版社

• 北京 •

图书在版编目(CIP)数据

大数据应用基础教程 / 刘三满主编. --北京:中国财政经济出版社,2025.2. -- ISBN 978-7-5223-3827-

9

I. TP274

中国国家版本馆 CIP 数据核字第 2025F7Q019 号

策划编辑:李昊民

责任校对:胡永立

责任编辑:张怡然

责任印制:张 健

大数据应用基础教程

DASHUJU YINGYONG JICHU JIAOCHENG

中国财政经济出版社 出版

URL: <http://www.cfeph.cn>

E-mail: cfeph@cfeng.cn

(版权所有 翻印必究)

社址:北京市海淀区阜成路甲 28 号 邮政编码:100142

营销中心电话:010-88191522

天猫网店:中国财政经济出版社旗舰店

网址: <https://zgczjjcbs.tmall.com>

廊坊市佳艺印务有限公司印刷 各地新华书店经销

成品尺寸:185mm×260mm 16 开 14 印张 300 000 字

2025 年 2 月第 1 版 2025 年 2 月河北第 1 次印刷

定价:49.00 元

ISBN 978-7-5223-3827-9

(图书出现印装问题,本社负责调换,电话:010-88190548)

本社质量投诉电话:010-88190744

打击盗版举报热线:010-88191661 QQ:2242791300

PREFACE

前言

在数字化、网络化、智能化日益普及的今天，大数据正以其独有的方式，渗透到各个行业和领域，深刻地影响着人们生活的方方面面。本教材顺应了时代发展的需要和产业发展的趋势(在党的二十大精神的指引下，依据《职业教育专业简介(2022年修订)》的最新要求进行编写)。对于学生来说，通过系统学习这门课程，掌握大数据处理、分析、挖掘等基础知识和基本技能，了解大数据在各行业中的应用案例和实践经验，无疑是提升职业竞争力和适应未来职场需求的关键，可以为未来的职业发展奠定坚实的基础。

本教材有以下特色。

一、思政融入，价值引领

本书将思政教育与专业知识紧密结合，以实际案例和讨论形式，将思政元素自然融入教学内容，使学生在掌握专业知识的同时，提升职业素养和社会责任感，实现知识传授与价值引领的有机统一。如通过讨论大数据、数字经济等前沿技术对社会、经济和生活的影响，引导学生关注科技进步与国家发展的关系，增强社会责任感和使命感；通过分析数字经济的诞生、发展及未来方向，激发学生对新兴技术的兴趣和对国家发展战略的认同，培养学生的创新精神和爱国情怀。



价值引领



体系完善



实践导向



业务场景驱动

二、知识体系完善，内容涵盖全面

本教材紧密围绕大数据应用的核心知识体系，从大数据的基本概念出发，逐步深入到大数据的采集、预处理、存储、管理、分析、挖掘、可视化以及安全保障等各个环节；内容全面而系统，循序渐进地引导学生构建起完整的大数据知识体系，为后续深入学习大数据技术奠定坚实的基础。

三、实践导向明确，注重应用能力培养

本教材秉持理论与实践相结合的教学理念，通过精心设计的实训案例和经过市场广泛验证的工具软件(如百度指数、WPS 文档等)，引导学生在实践中不断巩固所学知识，提升大数据应用的实际操作能力。这种教学方式不仅增强了学习的针对性和实用性，还有助于拓宽学生的大数据应用视野，提升他们的综合素养和创新能力。

四、业务场景驱动教学，工具软件精选实用

本教材并不拘泥于对某一工具软件的全面介绍，而是根据不同业务场景的实际需求，选择便捷、易用的工具软件来解决具体问题。这种业务场景驱动的教学方式，使学生能够更加直观地理解大数据技术在实际工作中的应用，同时也增强了学习的趣味性和互动性，激发了学生的学习兴趣和积极性。

本教材的编写出版，得到了山西警察学院院级规划教材立项的大力资助，同时还得益于公共安全类学科协作组 2024 年开放课题“公安信息技术应用”一流课程建设的理论探索和实践路径支持，教材的编纂还依托山西警察学院大数据与网络安全研究中心这一重要平台。

本教材不仅适合作为计算机专业与财经商贸类专业的必备教材，也是广大大数据爱好者的珍贵学习资源。在编纂过程中，我们有幸得到了众多院校领导的大力支持和参编教师的热情参与。他们的无私帮助和鼎力支持对本书的顺利完成发挥了巨大作用。在此，我们向所有给予我们帮助与支持的人表示最衷心的感谢。同时，我们也热切期盼广大读者能够提出宝贵的意见和建议，共同推动本书内容的不断完善和质量的持续提升，为大数据应用基础教育的发展贡献智慧与力量。

编 者

1 | 项目一 大数据概述

【学习目标】

【任务思维导图】

任务一 认知数据与大数据/2

任务二 探究大数据发展历程与应用
现状/6

任务三 认识数字经济/9

任务四 建立大数据思维/13

任务五 探索大数据相关技术与处理
流程/15

【课中实训】

实训 使用百度指数体验大数据
/23

【实训项目评价】

【课后习题】

31 | 项目二 大数据采集与预处理

【学习目标】

【任务思维导图】

任务一 大数据采集/32

任务二 大数据预处理的方法与
工具/41

【课中实训】

实训一 Python 与 PyCharm 的安装
配置/50

实训二 利用爬虫获取并解析数据/54

实训三 使用 Beautiful Soup 库
解析 HTML/56

实训四 数据清洗工具的安装与使用
/58

【实训项目评价】

【课后习题】

66 | 项目三 大数据存储与管理应用

【学习目标】

【任务思维导图】

任务一 认知数据存储/67

任务二 传统数据存储管理/70

任务三 大数据存储管理与工具/79

【课中实训】

实训一 安装和配置数据库/84

实训二 安装 Navicat Premium/88

实训三 创建与管理数据库/91

实训四 数据库导入/导出数据/94

【实训项目评价】

【课后习题】

102 | 项目四 大数据分析 with 数据挖掘

【学习目标】

【任务思维导图】

任务一 大数据分析/103

任务二 数据挖掘/108

任务三 机器学习/117

【课中实训】

实训一 安装与利用 NumPy 进行
排序/124

实训二 安装与利用 Pandas 进行统
计分析/127

实训三 安装与利用 Scikit-learn
进行分类/129

【实训项目评价】

【课后习题】

135 | 项目五 大数据可视化应用

【学习目标】

【任务思维导图】

任务一 数据可视化图表/136

任务二 可视化工具/141

【课中实训】

实训一 安装并利用 Matplotlib 绘制
柱状图/150

实训二 利用 Matplotlib 绘制饼图/152

实训三 安装并利用 Seaborn 绘制
散点图/154

【实训项目评价】

【课后习题】

165 | 项目六 大数据安全应用

【学习目标】

【任务思维导图】

任务一 大数据安全与威胁/166

任务二 大数据安全防护技术与
产品/176

【课中实训】

实训 WPS 文档的加密与解密/185

【实训项目评价】

【课后习题】

191 | 项目七 大数据的主要应用

【学习目标】

【任务思维导图】

任务一 大数据在国家安全领域的
应用/192

任务二 大数据在金融领域的
应用/193

任务三 大数据在健康医疗领域的
应用/202

任务四 大数据在城市管理领域的
应用/206

【课中实训】

实训 体验大数据的应用/211

【实训项目评价】

【课后习题】

参考文献/214

项目一

大数据概述



知识目标

- 1. 能够理解数据、信息与知识的关系，并提取和分析大数据中的有用信息。
- 2. 了解数字经济的催生因素、基本特征及其在现代经济中的地位。
- 3. 熟悉大数据技术的发展历程和当前趋势。

能力目标

- 1. 识别和分析大数据在不同领域的应用及其对业务的影响。
- 2. 能运用大数据思维分析和解决简单的问题。
- 3. 能独立分析和评估大数据相关信息。

素养目标

- 1. 关注大数据技术的发展趋势，培养创新精神。
- 2. 认识数据对经济发展的重要性，提升数据素养。

任务思维导图



任务一 认知数据与大数据



李明是一名电子商务专业的在校大学生，他最近对大数据产生了浓厚的兴趣，因为他意识到，在这个信息化飞速发展的时代，数据无处不在。李明回想起自己最近的购物经历，在某电商平台上，他仅仅浏览了几款心仪的商品，第二天，当他再次打开平台时，首页便推荐了许多与他之前浏览过的商品相似的产品。这种精准的个性化推荐，正是大数据技术的功劳。通过收集和分析用户的浏览、购买、搜索等行为数据，电商平台能够精准地预测用户的喜好和需求，从而提供更加个性化的服务。

李明深知，要想在这个大数据驱动的时代中脱颖而出，就必须对数据和大数据有深入的了解。于是，他决定从基础开始，一步步地探索数据的奥秘和大数据的神奇力量。

任务准备

一、数据、信息与知识

数据是信息的载体，它无处不在，形式多样。通过分析数据，我们可以提取有价值的信息，进而转化为知识，为决策提供支持。

（一）数据的含义

数据(Data)是客观事物的形式化表达，它涵盖了从原始的、未经加工的素材到多样化的表现形式。数据的范畴广泛，不仅涵盖数字，还包含文字、字母、符号的组合，以及图形、图像、视频、音频等多种媒介形式。它还能抽象地表示事物的属性、数量、位置及其相互关系。在日常生活中，我们周围充斥着各种数据。例如，在图书馆借阅的书籍记录、课堂参与度统计以及校园卡上的消费明细，这些看似纯数字的信息实际上都是数据的体现。此外，那些带有特殊符号的个人信息和验证机制，如××@××的校园邮箱地址，以及由字母和数字组成的课程验证码(如12AB)，同样是数据的一部分。

（二）数据的特征

数据显著的特征包括差异性、规律性和随机性。差异性体现在数据描述事物的独特属性和特征上，它使得每个数据集都有其独特的信息；规律性则表明数据分布遵循一定模式，这些模式可能揭示了数据背后的深层含义；而随机性则表现为数据中的不可预测性和不确定性，它使得数据分析充满挑战和机遇。数据分析的目标正是从看似杂乱无章、充满

随机性的数据中挖掘出隐藏的规律和联系，为企业、组织和个人提供有价值的决策支持。

（三）数据的存储单位

迅猛发展的互联网技术和智能设备的普及，实时记录与分析着人们的行为、位置和生理数据。数据存储单位不断扩展，从 Byte 至 KB、MB、GB、TB，直至 PB、EB、ZB、YB，乃至未来可能的 BB 级别，展现了数据规模的巨大增长。表 1-1 所示为各数据存储单位之间的换算关系。

表 1-1 数据存储单位之间的换算关系

单位	换算关系	单位	换算关系
Byte(字节)	1 Byte=8 BIT	PB(Petabyte, 拍字节)	1 PB=1 024 TB
KB(Kilobyte, 千字节)	1 KB=1 024 Byte	EB(Exabyte, 艾字节)	1 EB = 1 024 PB
MB(Megabyte, 兆字节)	1 MB=1 024 KB	ZB(Zettabyte, 泽字节)	1 ZB=1 024 EB
GB(Gigabyte, 吉字节)	1 GB=1 024 MB	YB(Yottabyte, 尧字节)	1 YB=1 024 ZB
TB(Terabyte, 太字节)	1 TB=1 024 GB	BB(Brontobyte, 布字节)	1 BB=1 024 YB

注意：NB(Nonabyte, 诺字节)和 DB(Doggabyte, 刀字节)并不是公认的存储单位标准，因此在表格中未列出。

（四）数据的分类

数据的分类可以从多个角度进行，以下是基于三个不同角度的分类。

1. 按性质不同分类

按性质不同，数据可以分为以下四种。

(1)定位数据。例如，GPS(Global Positioning System, 全球定位系统)坐标数据，用于精确记录地理位置。

(2)定性数据。例如，描述事物性质的分类数据，如城市名称、产品类型、动植物种类等。

(3)定量数据，包括具体的度量值，如建筑物的高度、货物的重量、车辆的速度等。

(4)定时数据。例如，日期、时间标记(或时间标签)，其用于记录事件发生的具体时间，如订单提交时间、会议开始时间等。

2. 按表现形式不同分类

按表现形式不同，数据可以分为以下两种。

(1)数字数据。这是各种统计或测量数据，它们以离散的形式存在，如整数、浮点数、符号、文字等，其用于精确表达数量、状态或标识。

(2)模拟数据。这类数据表现为连续的值，如声音波形、温度变化曲线等，它们能够连续地描述某种物理量的变化。

3. 按记录方式不同分类

按记录方式不同，数据可以分为以下三种。

(1) 普通文本格式数据。例如, TXT 文件(Text File, 文本文件)、CSV 文件(Comma-Separated Values File, 逗号分隔值文件)、Word 文档(Microsoft Word Document)等, 这些数据以人类可读的字符形式记录, 如文字、数字和标点符号。

(2) 二进制格式数据。这类数据以二进制编码的形式存储, 如 MP3(MPEG Audio Compression 3, 动态图像专家组音频压缩技术 3) 音频文件、MP4(MPEG-4 Part 14, 动态图像专家组标准第 14 部分) 视频文件、JPEG(Joint Photographic Experts Group, 联合图像专家组) 图片等。它们通常用于存储音频、视频、图像等多媒体信息。

(3) 混合数据。在某些情况下, 数据可能同时包含文本和二进制两种形式, 如某些复杂的文件格式(如 Word 文档中嵌入图片) 或数据库中的记录(包含文本描述和二进制附件)。在计算机系统中, 所有数据都以二进制 0 和 1 的形式表达和存储。

(五) 信息与知识

信息一般指数据经过处理和解释后所表达的含义或内容。知识则是指人们在社会实践中所获得的认识和经验的总和, 是经过加工、整理、归纳和提炼的信息。

数据、信息和知识的关系可概括为: 数据是基础的、记录事物属性的单元, 信息是对数据的解读和组织后的产物, 知识则是从信息中提炼出的具有普遍指导意义的规律。通过数据挖掘和分析, 我们可以从数据中提炼出有价值、新颖的模式, 进而转化为高层知识。

二、大数据

(一) 大数据的定义

大数据是指无法在一定时间内用常规软件工具对其内容进行抓取、管理和处理的数据集合。近年来, 随着技术的飞速发展, 大数据的普及程度、规模和价值均得到了显著提升, 其确切定义因使用场景不同而各异。

(二) 大数据的特征

大数据的“4V”特征: 大量性(Volume)、多样性(Variety)、高速性(Velocity)和价值性(Value), 具体如图 1-1 所示。

1. 大量性

大数据的一个重要特征就是其大量性, 表现在数据的容量大、体积大, 涉及数据的采集、存储和计算量都巨大, 从而构成了我们所说的海量数据。大数据的起始计量单位通常是 PB 级, 但实际中数据量之庞大往往远超这一级别, 可能达到 EB 或 ZB 级别。例如, 互联网公司日常运营中, 用户网络行为数据不断累积, 数据规模可能十分庞大。

2. 多样性

大数据的多样性主要体现在数据来源的多元化、数据类型的丰富化以及数据间的高度关联性。

(1) 数据来源的多元化。过去, 企业主要处理的是交易数据。然而, 随着互联网和物

联网的兴起，数据来源已扩展到社交网络、传感器等多个领域，形成了多样化的数据生态。

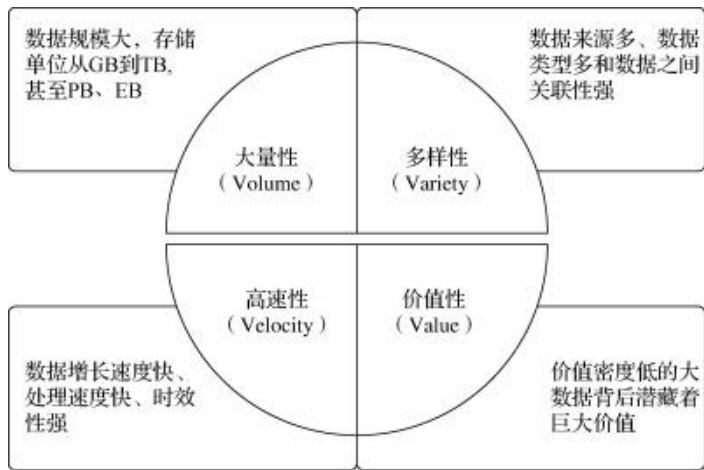


图 1-1 大数据的“4V”特征

(2) 数据类型的丰富化。大数据时代，数据格式日益丰富多样。目前，数据主要可分为三类：①结构化数据，如企业资源计划 (Enterprise Resource Planning, ERP) 系统、客户关系管理 (Customer Relationship Management, CRM) 系统等产生的数据，其特点是数据结构明确，逻辑关系清晰。②半结构化数据，如超文本标记语言 (Hypertext Markup Language, HTML) 文档、电子邮件等，其结构不像数据库中的表结构那么严格，但具有一定的组织结构。③非结构化数据，如社交媒体上的文本、图片、视频等，这些数据往往没有固定的数据结构，难以用传统的数据库系统进行处理。

目前，70%~85%的数据以半结构化数据和非结构化数据的形式存在，而传统的结构化数据只占较小的比重。

(3) 数据间的高度关联性。大数据的多样性不仅体现在数据来源和类型的丰富性上，还显著表现在数据之间的紧密关联性上。例如，学生在旅行过程中分享的照片和日志，不仅记录了他们的旅行经历，还与其所在位置、行程安排等信息紧密相关。

3. 高速性

高速性是指大数据具有快速增长、高效处理以及强时效性的特点，这是大数据与传统数据显著区别的重要标志。在高速性的要求下，数据必须能够快速被抓取、处理和响应。例如，社交媒体平台上，用户发布的实时动态需要被即时捕捉并展示给关注者；在金融领域，高频交易系统需要毫秒级的数据处理速度来确保交易的实时性和准确性；在交通管理系统中，交通流量和路况信息的实时更新有助于驾驶员做出更明智的出行决策。

4. 价值性

大数据虽然庞大，但真正能够转化为实际价值的内容往往只占其中一小部分。尽管如此，这些价值密度低的数据背后却蕴藏着巨大的潜在价值。大数据的真正价值在于，运用

机器学习、人工智能及数据挖掘技术，从庞杂、多样化的数据海洋中提炼出对未来趋势和模式预测至关重要的信息。这些被提炼出来的知识可以应用于农业、金融、医疗等多个领域，为决策提供支持，从而创造更大的价值。



请思考并分组讨论大数据对大学生日常生活的影响。

任务二 探究大数据发展历程与应用现状



李明是一名电子商务专业的在校大学生，在一次课程讨论中，他听到老师提及大数据技术在现代商业中的广泛应用及其对决策制定的深远影响。老师提到，大数据技术不仅能够帮助企业深度洞察消费者行为，还能优化运营策略，提升市场竞争力。

李明对此产生了浓厚的兴趣，因为他意识到，作为一名电子商务专业的学生，了解并掌握大数据技术将是未来职业生涯中的一大优势。然而，当他在尝试进一步了解大数据技术的发展历程和应用现状时，却发现自己对这一领域的了解非常有限。

任务准备

一、 大数据发展历程

（一）第一阶段：萌芽期（20 世纪 90 年代至 21 世纪初）

大数据的概念起源于美国，早在 1980 年，著名未来学家阿尔文·托夫勒 (Alvin Toffler) 便在《第三次浪潮》一书中预见性地提及大数据，将其誉为第三次浪潮中的璀璨篇章。1997 年，NASA 阿姆斯科研中心 (NASA Ames Research Center) 的大卫·埃尔斯沃斯 (David Ellsworth) 和迈克尔·考克斯 (Michael Cox) 在研究数据可视化时，正式提出了“大数据”这一术语。

这一阶段，数据挖掘理论和数据库技术逐渐完善，商业智能工具和知识管理技术如数据仓库、专家系统、知识管理系统等开始在企业、机构中广泛应用，助力其对内部数据进行深度统计、分析、挖掘和利用，从而优化决策过程。

（二）第二阶段：成熟期（21 世纪初至 2010 年）

在 21 世纪的前十年，互联网行业蓬勃发展，IT 技术不断革新，其中大数据技术率先在互联网行业中崭露头角。2008 年，《自然》杂志以“大数据”为主题推出封面专栏，进一

步提升了公众对大数据的关注度。

这一阶段，Web 2.0 应用的崛起使非结构化数据激增，传统处理方式显得捉襟见肘，这催生了大数据技术的迅速进步。大数据解决方案逐渐完善，并行计算和分布式系统成为支撑其发展的两大核心。电信、银行、电商等海量数据行业开始运用这些技术，对社会经济的多个领域产生了广泛而深远的影响。

（三）第三阶段：大规模应用期（2010 年后）

自 2010 年起，全球多国和地区竞相制定大数据战略。2012 年，瑞士达沃斯世界经济论坛聚焦大数据，并发布《大数据，大影响》(Big Data, Big Impact) 报告，正式将大数据确立为新型经济资产。2013 年，标志着“大数据元年”的来临，其技术迅速渗透到商业、科技、医疗、政府、教育、经济、交通、物流等多个领域。自 2014 年起，大数据技术迎来了持续发展和广泛应用的黄金时期。各国企业纷纷增加投资，以推动大数据技术的研发和应用。在数据收集、存储、处理和分析等方面，大数据技术取得了显著的突破，数据处理能力和速度得到了极大的提升。

在这一阶段，大数据应用不再局限于特定行业，而是实现了跨行业、跨领域的数据整合。企业和社会组织积极寻求整合不同来源、结构的数据，以挖掘数据间的关联性和潜在价值，为社会治理和产业发展提供更为全面、深入的数据支撑。

二、大数据的应用现状

大数据产生于互联网领域，并逐步推广到电信、金融、交通等领域，在众多行业中产生了实用价值。

（一）互联网领域

互联网企业凭借大数据的运用，能够深度洞察客户行为，从而制定出高度个性化的服务策略，显著提升企业效益。实践证明，有效利用大数据技术能够将电子商务的营业效率大幅提升。

电商平台通过大数据收集用户的多样化数据，并利用这些数据构建详尽的“用户画像”，进而实现个性化推荐、精准营销和广告投放。举例来说，当用户登录电商网站时，系统能够依据“用户画像”精准预测用户可能的购买需求。例如，如果用户过去经常购买户外用品，系统可能会在首页推荐最新款的徒步鞋或户外背包，并以“猜你喜欢”的方式展示给用户。再如，如果用户的购物车中有多包羊肉片和糖蒜，但缺少火锅蘸料，系统会在结账时智能提示是否需要添加蘸料，并在用户确认后自动引导至蘸料商品页面，这极大提升了用户购物体验和购买转化率。

在校园生活中，数据也发挥着同样重要的作用。举例来说，当学生在使用校园内的在线学习平台时，系统能够依据“学习画像”来精准预测学生的学习需求。如果学生的学习记录显示他们经常查阅数学课程的资料和习题，学习平台可能会在首页推荐数学的高级课程

或相关的学习资料，并以“你可能感兴趣”的方式展示给学生。另外，如果学生在选课系统中选择了多门与编程相关的课程，但尚未报名参与编程竞赛，系统可能会在选课结束后智能提示学生是否对这些活动感兴趣，并在学生确认后自动引导至相关的活动页面，帮助他们更好地规划和扩展学习体验。通过这种智能化的推荐和提示，系统能够极大提升学生的学习效率和体验，帮助他们更好地学习和成长。

（二）电信领域

在电信行业中，用户日常产生的语音、短信、流量和宽带数据构成了庞大的数据资源。借助大数据技术，运营商能够显著提升数据处理能力，实现海量数据的聚合与深度洞察。目前，电信领域主要将大数据应用于以下 5 个方面。

（1）网络管理与优化。通过大数据分析，运营商可以对基站选址等基础设施建设进行精准优化，确保网络覆盖更加合理；同时，对已有设施进行效率和成本评估，减少资源浪费，提升网络性能。

（2）市场与精准营销。利用大数据技术进行客户画像描绘、关系链研究，以及实现精准营销、实时营销和个性化推荐。这种精准化的营销策略能够大幅提高营销效率，提升客户满意度。

（3）客户关系管理，包括客服中心优化和客户生命周期管理。通过对用户数据的深入分析，运营商能够及时发现潜在的“离网”客户，并据此制定个性化的挽留措施，防止客户流失，保持市场份额的稳定。

（4）企业运营管理。基于企业内部的业务和用户数据，以及通过大数据手段收集的外部社交网络数据、技术和市场数据，运营商可以对业务和市场经营状况进行全面、深入的分析，为企业的战略决策提供有力支持。

（5）数据商业化与增值。运营商积极寻求与第三方合作伙伴的联动，以最大化数据的商业价值。这主要体现在两个方面：数据即服务（Data as a Service, DaaS）和分析即服务（Analytics as a Service, AaaS）。一方面，通过 DaaS 模式，运营商开放脱敏后的数据或 API，为外部合作伙伴提供有价值的数据资源。另一方面，借助 AaaS 模式，运营商与第三方公司携手，利用脱敏数据为政府、企业和行业客户提供定制化的信息、数据建模和深度分析服务，实现数据价值的最大化与外部变现。

（三）金融领域

银行在多年的数据积累基础上，可积极探索利用大数据来推动业务运营的全面革新。其大数据应用主要涵盖以下 4 个方面。

（1）客户画像的精准应用。客户画像的精准应用涵盖个人与企业两大核心领域。个人客户画像侧重于捕捉人口统计学特征、消费能力、兴趣偏好及风险承受能力等关键信息；而企业客户画像则更为全面，整合了企业的生产、流通、运营、财务、销售数据，以及客户与产业链上下游的全方位信息，为企业提供更深入的市场洞察和决策支持。

（2）精准营销。基于详尽的客户画像，银行能够实施精准营销策略。通过深度分析客

户的年龄、资产规模、理财偏好等特征，银行能够精准定位目标客群，挖掘其潜在的金融服务需求，并据此推出个性化的服务与产品。

(3) 风险管控的智能化。大数据在风险管控中发挥着关键作用，特别是在中小企业贷款风险评估和欺诈交易识别方面。银行通过整合持卡人信息、交易历史、行为模式等数据，结合智能分析技术，实现了对交易行为的实时反欺诈监测，显著提升了风险识别和防控能力。

(4) 运营优化升级。运营优化涉及市场渠道和产品服务的全面升级。借助大数据，银行能够实时追踪并评估不同市场推广渠道的效果，特别是网络渠道的推广质量，从而精准调整和优化合作渠道策略。同时，银行还能将客户行为数据转化为信息流，深入分析客户的个性特征、风险偏好和消费习惯，以更深入地理解客户需求，进而实现智能化的产品创新和服务优化。

(四) 交通领域

在交通领域，大数据技术的应用正在推动智能交通系统的快速发展。通过对交通流量、车辆行驶轨迹等数据的收集和分析，交通管理部门可以实现对交通拥堵的预警和疏导，优化交通资源配置，提高道路通行效率。同时，大数据技术还可以帮助交通管理部门实现对交通违法行为的快速识别和打击，保障道路交通的安全和顺畅。此外，大数据技术还可以为智能交通系统提供数据支持，实现车辆导航、智能停车等功能，提升人们的出行体验。

此外，大数据在社会治理、体育娱乐等领域也展现出了广泛的应用前景。展望未来，大数据将在更多领域深化应用，推动社会进步，造福广大民众。



分组讨论大数据技术未来的发展方向和可能面临的挑战。

任务三 认识数字经济



李明是一名电子商务专业的在校大学生，他在学习过程中逐渐认识到数字经济正成为全球经济发展的重要趋势。一次，在翻阅关于新兴技术的资料时，他注意到了“数字经济”这一概念，并了解到数字经济以数据资源为核心，借助大数据、云计算、物联网等前沿技术，正在深刻改变人们的生产方式、生活方式和社会治理方式。

李明意识到，作为电子商务专业的学生，了解并掌握数字经济的诞生、基本特征以及建设方向，对他未来的学习和职业发展至关重要。

 任务准备

一、数字经济的诞生

大数据，作为信息技术发展的产物，开启了信息化新阶段，从辅助工具转变为引领社会经济发展的核心引擎，催生了“数字经济”这一全新经济形态。

数字经济通过对大数据的识别、筛选、存储与应用，实现资源的高效配置与再生，推动经济高质量发展。数字经济涵盖直接或间接利用数据引导资源、推动生产力发展的各种经济形态。在技术层面，数字经济融合大数据、云计算、物联网、人工智能和 5G 通信等前沿技术；在应用层面，新零售、新制造等成为其典型代表。

数字经济以数据资源为核心，依托现代信息网络，借助信息通信技术的融合应用和全要素数字化转型，追求公平与效率的和谐统一。其正在深刻改变生产方式、生活方式和治理方式，成为全球要素资源重组、经济结构重塑、竞争格局改变的关键力量。

二、数字经济的基本特征

数字经济受三大核心定律的深刻影响。首先，梅特卡夫法则揭示了网络价值的本质，其价值随着用户数平方的增长而增长，即与用户数平方成正比，联网计算机的增加会呈指数级地提升整体网络及其每个参与者的价值。其次，摩尔定律描绘了信息技术的演进轨迹，即计算机硅芯片的处理能力每 18~24 个月翻倍，而成本减半，体现了技术进步的高速和成本的显著降低。再者，达维多定律则强调了市场先入者的重要性及其对产品迭代的紧迫性。尽管这一定律不直接要求企业淘汰旧产品，但它凸显了在市场竞争中，持续创新和快速迭代对于保持领先地位的关键作用。尽管达维多定律与马太效应在优势累积方面有相似之处，但两者在本质和应用上仍有所区别。

这三大定律共同塑造了数字经济的基本特征：快捷性与高渗透性、自我膨胀性与边际收益递增性、共享增效与绿色可持续发展、高效直联性。

（一）快捷性与高渗透性

互联网以其无与伦比的连接能力，打破了地域界限，将全球紧密联结成一个“地球村”。它消解了时间的束缚，使得信息传输和经济活动得以在极短的时间内完成，因此，数字经济展现出了显著的快捷性特征。现代信息网络更是以光速传递信息，使得数字经济在信息的收集、处理和应用上接近实时，形成了一种以速度为驱动的经济形态。

与此同时，信息技术及网络技术的迅猛发展，使数字经济具备了极强的渗透性。这种渗透性不仅迅速深入到了第一产业、第二产业和第三产业，还使得这些产业之间的界限日益模糊，促进了产业间的深度融合与相互渗透。这种高渗透性不仅推动了传统产业的数字

化转型，更催生了众多基于互联网的新兴产业，为全球经济的多元化和动态化注入了新的活力。

（二）自我膨胀性与边际收益递增性

数字经济的独特价值源于其网络节点数的平方效应，这意味着随着网络用户的增长，其产生的效益将呈现指数级跃升。在数字经济生态中，由于网络效应和用户行为的集中趋势，一旦某个产品或服务获得了初步的市场优势，这种优势将在特定条件下持续自我强化，形成“强者恒强，弱者愈弱”的竞争格局，最终导致“赢家通吃”的市场格局。

边际收益递增性在数字经济中尤为显著。在知识密集型经济背景下，随着知识与技术要素的持续投入，生产者的收益将呈现显著递增趋势。这种经济规律对企业经营具有深远影响。在数字经济中，边际收益递增性主要体现在两个方面。

首先，数字经济的边际成本递减特性明显。由于数字产品的可复制性和共享性，随着生产规模的扩大，单位产品的成本会不断降低，即边际成本递减。这使得数字经济中的企业能够以较低的成本提供大量产品和服务，实现规模经济效应。

其次，数字经济展现出累积增值性。在数字经济中，信息和知识的价值具有累积性。随着数据的不断积累和分析，企业能够更精准地把握市场需求和消费者行为，提供更为个性化的产品和服务，从而增加收益。同时，这种累积增值性也体现在企业的无形资产上，如品牌、专利、用户数据等，这些资产能够为企业带来持续的价值增长。

（三）共享增效与绿色可持续发展

（1）共享增效。在数字经济中，产品效用与用户数量呈正比增长。网络效应和资源共享为用户带来更高的便利性和效率，使得用户满意度和忠诚度随之增强。这种共享增效不仅提升了用户体验，更为数字经济的蓬勃发展注入了强劲动力。

（2）绿色可持续发展。数字经济以其独特的方式，有效改善了传统工业生产对资源、能源的过度消耗以及由此产生的环境问题的状况。通过数字化、智能化技术的应用，数字经济显著提高了资源利用效率，减少了浪费和污染，为社会的可持续发展贡献了重要力量。同时，它也为绿色产业的兴起提供了强大支持，成为推动环境保护和可持续发展的重要引擎。

（四）高效直连性

随着网络技术的突飞猛进，经济组织结构正迈向扁平化。在这一趋势中，生产者与消费者能够通过网络直接相连，实现无缝交易，从而大幅减少了中间环节的需求。这种高效直连性不仅显著降低了交易成本（缩减了中间环节的费用和耗时），更提升了经济效益。生产者与消费者能够更精准地满足对方需求，减少了信息不对称和误解，为经济活动带来了更高的效率和更低的成本，成为数字经济的一大鲜明特色。

三、数字经济的建设方向

（一）核心技术自主创新

我国应聚焦数字关键核心技术，发挥制度与市场优势，强化基础研发能力。通过实施核心技术攻关工程，突破发展瓶颈，实现高水平自立自强，确保数字经济自主权。

（二）新型信息基础设施建设

要加快新型信息基础设施的布局和建设，重点围绕 5G 网络、一体化数据中心和国家产业互联网等。构建智能化、综合性、安全可控的数字信息基础设施体系，打通经济社会发展的信息通道。同时，推动产业化、规模化应用，培育具有国际竞争力的软件企业，突破关键软件技术，提升软件产业的创新能力和供给能力。

（三）深化数字经济与实体经济的融合

顺应数字化、网络化、智能化潮流，我们应积极推进制造业、服务业、农业的数字化转型。利用互联网新技术对传统产业进行全面革新，提升生产效率，发挥数字技术对经济的倍增作用。推动互联网、大数据、人工智能与产业的深度融合，培育行业领军企业，确保数字化转型与企业战略相契合，避免盲目性。

（四）聚焦重点领域，发展数字产业

我们需聚焦于战略前沿与关键领域，基于技术突破和市场需求，提升产业链关键环节竞争力，完善供应链体系，加速产品和服务创新。特别关注集成电路、新型显示、通信设备、智能硬件等领域，加大研发投入，培育具备国际竞争力的大企业。同时，通过政策引导与资源整合，推动数字产业集群化发展，打造世界级数字产业集群，增强我国在全球数字经济领域的竞争力。

（五）规范数字经济发展

数字经济发展必须与发展监管并重。完善市场准入、公平竞争审查与监管机制，构建全方位监管体系，确保全链条、全领域的高效监管。严厉打击损害群众利益、妨碍公平竞争的行为，防范平台垄断和资本无序扩张，维护公平竞争环境。同时，保护从业人员和消费者权益，强化税收监管，确保税收公平有效。

（六）构建现代化数字经济治理体系

推动数字经济治理现代化，需健全法律法规和政策体系。明确主管部门与监管机构的职责，强化部门间协同。提升监管技术，将监管融入创新、生产、经营、投资各环节。强化平台企业主体责任，建立行业自律机制，鼓励企业守法守规。同时，引入社会、媒体和公众监督，形成共治合力。加强数字经济安全风险防控，确保关键领域安全可控。深化理论研究，为数字经济发展提供理论支持。

（七）深化数字经济国际合作

积极参与国际数字经济合作，关注国际动态，参与数字经济议题谈判，推动多边数字治理合作。在此过程中，积极贡献中国方案，展现中国智慧与担当，为构建开放、包容、普惠、平衡、共赢的全球数字经济治理体系贡献力量。



请讨论数据作为数字经济时代的新生产要素，如何影响传统经济形态下的资源分配和生产方式。

任务四 建立大数据思维



李明是电子商务专业的在校大学生。在一次校园电商平台的促销活动策划中，他面临如何精准推送优惠信息以提高学生参与度的难题。他意识到传统思维可能存在不足，于是产生了要运用大数据思维的想法。

任务准备

从传统的机械思维到现代的大数据思维，每一次思维模式的演进都极大地推动了人类社会的进步。机械思维，以其确定性、可描述性和普适性，曾在科学、工程和经济学等领域发挥重要作用。然而，机械思维在面对复杂、多变和不确定性的现实世界时，逐渐显露出其局限性。它往往过于依赖预设的模型和假设，难以应对海量、高速和多样化的数据。大数据思维则是一种基于数据洞察世界的全新范式，它强调量化分析、重视事实依据、不懈追求真理。大数据研究专家舍恩伯格(Schonberger)指出，大数据时代，人们对待数据的思维方式会发生以下变化：第一，人们处理的数据从样本数据变成全部数据；第二，由于是全样本数据，人们不得不接受数据的混杂性，而放弃对精确性的追求；第三，人类通过对大数据的处理，放弃对因果关系的渴求，转而关注相关关系。当然，大数据时代带给人们思维方式的转变可能不止上述三个方面。大数据思维更大的转变在于建立以数据为核心的智能思维方式，让数据创造价值，并赋予类似于“人脑”的智能，甚至智慧。大数据思维体系如图1-2所示。

1. 总体性思维：全样而非抽样

全样而非抽样指的是人们处理的数据从样本数据变成全部数据。在大数据时代，人们能够获取并分析海量的数据，甚至是相关的所有数据，摒弃了传统上对采样的依赖。这种变革使得我们能够对事物有更全面的认识，能够揭示样本无法提供的细微信息和深层规

律。随着数据采集、处理和分析技术的飞速进步，我们已能高效、动态地获取与研究对象相关的全面数据。这些技术突破让我们超越了样本研究的局限，实现了从样本到总体的思维转变。如今，我们能够以更加直观、全面、立体和系统的方式洞察总体状况，为决策和预测提供更准确的依据。

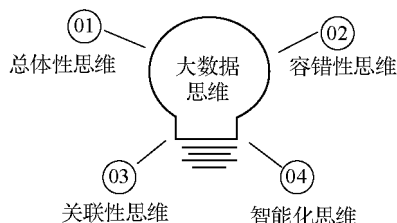


图 1-2 大数据思维体系

2. 容错性思维：效率而非精确

效率而非精确指的是由于面对的是海量数据和全样本数据，因此人们不得不接受数据的混杂性，而放弃对精确性的追求。在数据稀缺的时代，精确性对于数据分析至关重要，因为它直接影响结论的准确性。然而，在大数据时代，海量的即时数据使得我们不再需要过分追求每个数据点的精确性。相反，我们应采取容错性思维，即在一定程度上容忍数据的错误和混杂。这种转变允许我们更专注于宏观层面的洞察，而非拘泥于微观层面的精确。通过这种方式，我们可以利用大数据的优势，获得更全面、更深入的认知。

3. 关联性思维：相关而非因果

相关而非因果指的是人们通过对大数据的处理，放弃对因果关系的渴求，转而关注相关关系。在大数据时代之前，人们往往追求现象背后的因果关系，试图通过有限的样本数据解析其内在联系。然而，小数据量难以全面反映事物间的普遍关联。在大数据时代，数据挖掘技术让我们能够揭示事物间隐藏的关联，获取更深层次的认知。这促使我们转变思维方式，从因果思维转向相关思维，摒弃传统偏见，以全面洞察大数据带来的新机遇，更准确地把握现在并预测未来。

4. 智能化思维：以数据为中心

以数据为中心指的是在决策、分析、创新以及业务运营的各个方面，将数据作为核心和出发点。尽管信息社会已显著提升自动化和智能化水平，但机械思维仍受限于线性、简单的自然模式。大数据的兴起为机器智能的飞跃提供了可能。通过机器学习，机器能从海量数据中提炼出有价值的信息，实现由自然思维向智能化思维的转变，这是大数据思维的核心转变。随着物联网、云计算和可视化技术的进步，大数据系统智能地搜索、分析数据，提供主动、立体的洞察。智能与智慧成为大数据时代的核心特征。

大数据的发展是数据、技术与思维三大要素共同作用的结果。它不仅仅依赖于大数据资源的不断扩展和大数据技术的持续创新，更关键的是大数据思维的形成。没有大数据思维作为指引，即便拥有海量的数据资源，也难以充分发挥其潜在价值。因此，数据、技术

与思维三者缺一不可，它们相互促进，共同推动大社会的繁荣发展。



分组讨论个人在日常生活和工作中如何运用大数据思维来做出更明智的决策。

任务五 探索大数据相关技术与处理流程



大数据不仅仅是海量数据的集合，更是一种全新的思维方式和技术手段，它能够帮助我们更深入地理解世界，更准确地预测未来。对于电子商务专业的学生李明而言，了解大数据相关技术与处理流程，对于实现数据的价值转化、提升业务效率和优化用户体验至关重要。



任务准备

一、 大数据应用技术

（一）云计算

1. 云计算的概念

云计算是一种通过网络将计算任务分配到多个服务器进行分布式处理的计算模式。它整合了计算、存储和服务资源，通过虚拟化技术为用户提供灵活、可扩展的服务。

2. 云计算的基本原理

云计算的核心原理是数据和服务资源的集中化管理和用户按需访问。用户数据存储在远程数据中心，并由云服务提供商负责维护。用户可通过互联网设备随时随地访问所需资源，如存储、数据库、服务器和应用软件，并按使用量付费，实现高效、便捷的计算资源使用。

3. 云计算的分类

云计算可分为多种类型，以满足不同需求。

(1) 根据服务模式，云计算分为三种。

①基础设施即服务 (Infrastructure as a Service, IaaS)，提供虚拟化硬件资源，如服务器、存储和网络，客户可以在这些资源上部署自己的软件。

②平台即服务 (Platform as a Service, PaaS)，提供完整的开发和部署环境，包括操作系统、数据库、中间件等，客户可以开发、测试和运行自己的应用程序。

③软件即服务(Software as a Service, SaaS), 提供预装的软件应用程序, 客户无须安装或维护软件, 即可通过互联网使用。

(2) 根据部署方式, 云计算分为四种。

①公有云, 是由第三方供应商提供的云服务, 通过互联网访问, 如阿里云、腾讯云等。

②私有云, 为单一组织构建的云服务, 由组织内部拥有和管理, 以满足特定的数据、安全性和服务质量需求。

③社区云, 是由多个具有相似需求的组织共同拥有的云服务, 通常用于共享基础设施, 以实现成本降低和信息共享。

④混合云, 结合了公有云、私有云和社区云中的两种或多种部署方式, 为企业提供更多的灵活性和更多数据存储选项。

4. 云计算与大数据的关系

云计算与大数据之间的关系非常密切, 它们相互依存、相互促进, 共同推动了大数据的广泛应用和发展。

(1) 云计算是大数据的基础。云计算通过提供弹性扩展的计算和存储资源, 满足了大数据处理对庞大资源的需求。其虚拟化技术使得硬件资源得以高效利用, 降低了数据处理成本, 使得大数据处理更加灵活、高效。

(2) 大数据是云计算的重要应用。云计算平台为大数据处理提供了高效、可靠的环境。在云端, 大数据可以进行快速收集、存储、分析和应用, 使得数据的价值得以最大化。同时, 云计算提供的丰富数据处理工具和服务, 进一步推动了大数据的广泛应用。

(3) 云计算与大数据相互促进。云计算不断优化其服务质量和资源分配效率, 为大数据处理提供了更好的支持。同时, 大数据处理的需求也推动了云计算技术的不断发展和创新。两者之间的紧密合作, 共同推动了技术的进步和应用的拓展。

总之, 云计算与大数据之间的关系紧密。在未来, 随着技术的不断进步和应用场景的不断拓展, 云计算与大数据将继续发挥重要作用, 推动数字化转型和智能化升级。

(二) 物联网

1. 物联网的概念

物联网(Internet of Things, IoT)是新一代信息技术的重要组成部分, 意指物物相连, 万物万联。物联网描述了一个全新的网络形态, 它通过将各种物理设备、传感器、执行器等连接到互联网, 实现这些物品之间的智能化识别、定位、跟踪、监控和管理。

2. 物联网的关键应用技术

①传感器应用技术: 用于收集物品的状态信息, 如温度、湿度、压力等。②RFID 标签(Radio Frequency Identification Tag, 无线射频识别标签): 通过无线电信号识别特定目标并读写相关数据, 无须建立机械或光学接触。③嵌入式系统应用技术: 将计算机硬件、软件和相关算法集成到一个单一设备中, 实现设备的智能化和自动化。

物联网应用技术正在逐步改变我们的生活方式和工作模式，它使得各种物品能够相互连接、相互通信，实现智能化管理和控制。在智能家居、智能交通、工业自动化等领域，物联网技术已经得到了广泛应用。

3. 物联网的体系架构

物联网的体系架构可以分为三层，自下而上依次是感知层、网络层、应用层，如图 1-3 所示。

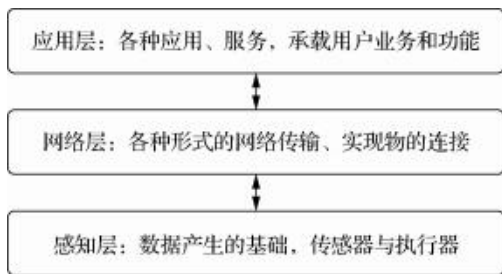


图 1-3 物联网的体系架构

(1) 感知层。感知层是物联网的感知单元，负责识别和采集各种物体的信息。它包含信息采集子层和通信子层。以传感器、射频识别(Radio Frequency Identification, RFID)、智能装置、二维码、条形码等技术为数据采集工具，并通过通信子层的通信模块和延伸网络，将采集到的数据发送到网络层，与网络层的网关进行信息交换。

(2) 网络层。网络层是物联网的通信核心，负责将感知层收集的信息传输到云端或其他处理节点。它类似于一个高效的信息传输系统，确保数据的快速、准确传输。网络层的主要任务是处理来自感知层的信息，并根据需要决定是将这些信息发送到云端进一步处理，还是直接进行其他操作。

(3) 应用层。应用层是物联网的智能处理中心，位于云端的主机将接收来自网络层的大量数据。经过大数据分析后，应用层能够做出智能决策，并向各个装置发送相应的指令，以实现监测与控制的自动化和智能化。

4. 物联网与大数据的关系

物联网通过其感知层技术使物体与物体在网络中互联互通，实现物物感知、信息共享、协作联通、智慧应用，从而使得物体具备一定的“智能”。物联网不仅仅赋予物体“智慧”，还可以通过数据分析和处理提高这种“智能”水平。物联网确实需要借助大数据技术来实现对物联网数据的智能分析和处理。

大数据的发展确实得益于物联网技术的广泛应用。物联网中的传感器和智能设备持续不断地产生大量数据，这些数据构成了大数据的主要来源之一。没有物联网技术的快速发展，数据产生的方式就不会发生这样的变革，即由人工产生为主转向自动产生为主，大数据时代也不会如此迅速地到来。因此，物联网是大数据技术的关键数据源之一。

(三) 移动互联网

1. 移动互联网的概念

移动互联网是移动通信终端与互联网的结合，用户可通过智能手机、平板电脑等无线设备，在移动(如在火车上、步行时、自驾车辆中等)中借助高效能的移动网络，无缝地接入网络，享受多样化的网络服务。除了传统应用，如网页浏览和文件下载，增强现实(Augmented Reality, AR)、虚拟现实(Virtual Reality, VR)、实时视频、智能家居远程操控以及基于人工智能(Artificial Intelligence, AI)的个性化推荐等应用也逐渐成为主流。这些新兴应用不仅丰富了用户体验，也推动了移动互联网技术的持续进步和创新。

2. 移动互联网的组成

移动互联网可以划分为移动互联终端设备、移动通信网络、移动互联网应用和移动互联网相关应用技术四大部分，如图 1-4 所示。

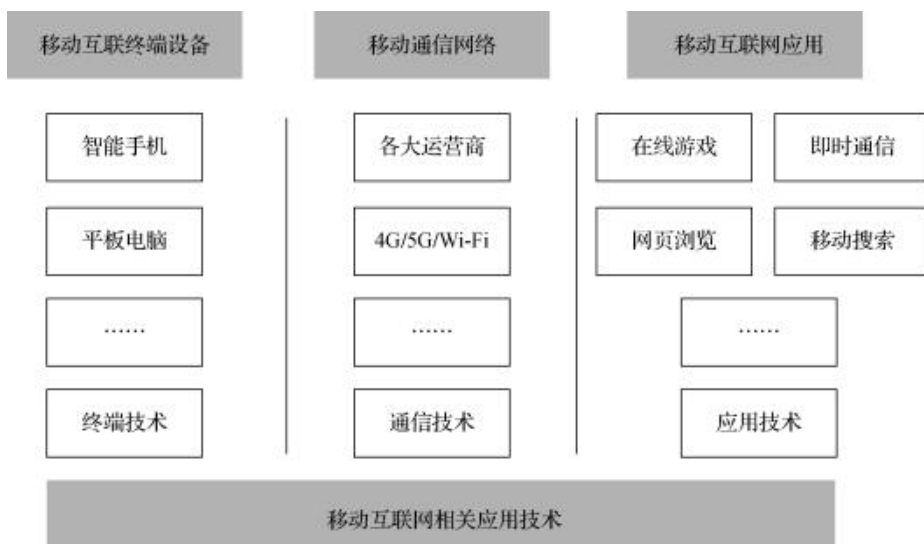


图 1-4 移动互联网结构

(1) 移动互联终端设备。移动互联终端设备是推动移动互联网发展的关键力量。这些设备，如智能手机、平板电脑等，通过集成无线网络技术，使用户能够随时随地接入互联网，享受各种服务，从而极大地推动了移动互联网的快速发展。

(2) 移动通信网络。移动通信网络是移动互联网的基础设施。通过无线网络技术，移动通信网络能够覆盖广泛的地域，使用户能够在任何地点接入互联网。这为移动互联网的普及和应用奠定了坚实的基础。

(3) 移动互联网应用。移动互联网应用是移动互联网的核心。各种类型的应用程序，如社交媒体、在线购物、移动支付等，已经渗透到人们的日常生活和工作中，极大地丰富了移动互联网的用户体验，并推动了移动互联网的持续发展。

(4) 移动互联网相关应用技术。移动互联网的发展离不开相关技术的支持。这些技术包括终端技术、通信技术和应用技术。终端技术主要涉及硬件设计和智能操作系统开发；通信技术涵盖通信标准、协议、移动网络技术和短距离无线通信技术；应用技术包括服务器端技术、浏览器技术和移动互联网安全技术。这些技术的不断进步和创新，为移动互联网的发展提供了强大的技术支撑。

3. 移动互联网与大数据的关系

移动互联网与大数据紧密相联。一方面，智能手机、智能穿戴设备及物联网技术的普及，为大数据提供了海量的实时数据，包括用户的日常生活、消费偏好和地理位置等，极大地丰富了大数据的素材来源。另一方面，移动互联网应用正逐步融入大数据技术，为用户提供更个性化的服务，如电商平台的商品推荐和社交媒体的内容推送。这些应用基于大数据驱动，推动了移动互联网的进一步发展。同时，大数据处理能力的提升使得数据更新更加迅速，无论用户身处何地，都能实时收集、传输和处理数据，使大数据在移动互联网领域的应用更加广泛和深入。

（四）人工智能

1. 人工智能的概念

人工智能，作为一门新兴技术科学，专注于模拟、扩展人类智能的理论、方法、技术及应用系统。其起源可追溯至 20 世纪 50 年代，当时达特茅斯学院的数学助理教授约翰·麦卡锡聚集了马文·明斯基、克劳德·香农等多位科学家，共同探讨机器模仿人类智能的可能性。尽管会议未达成共识，但“人工智能”这一术语由此诞生，1956 年也因此被誉为人工智能元年。

人工智能涵盖机器学习、计算机视觉、机器人技术等多个研究领域，能模拟人类的意识和思维过程。但必须明确，人工智能并非真正的人类智能，而是模拟人类智能的计算机系统。在特定任务中，其甚至能超越人类智能。

人工智能的核心在于通过大量样本数据和算法引导，使机器能自动识别相似目标。这需要人工参与数据标注和模型训练，以实现机器的智能识别。若面对未经训练的新物品，计算机通常无法识别。

2. 人工智能的研究成果应用

(1) 问题求解。该功能包括问题表示、搜索算法和行动计划等，它是在人工智能早期研究的基础上逐渐发展起来的，包括棋类游戏、智力问答、定理证明等。问题求解技术给人们留下的深刻印象之一就是人机对弈，人工智能棋手曾经在围棋大赛上战胜了许多人类高手。

(2) 模式识别。利用智能设备代替人类或者辅助人类感知外部信息，如指纹识别、人像识别、文字识别、图像识别、车牌识别、语音识别等。这些技术已经广泛应用于安全认证、智能监控、智能输入等领域。

(3) 自动化工程。主要应用领域包括自动驾驶、自动化生产线、智能家居等。这些应用通过智能系统实现对设备的自动化控制和管理,提高了生产效率和生活质量。

(4) 知识工程:专注于利用人工智能和软件技术设计、构建和维护知识系统。该系统范畴涵盖专家系统、智能搜索、计算机视觉、机器翻译、自然语言处理、深度学习及数据挖掘等。这些技术已广泛应用于医疗、金融、教育等领域,为用户提供智能化、高效化的服务体验。

3. 人工智能与大数据的关系

人工智能与大数据紧密相联,协同工作。人工智能如同渴求知识的学者,依赖大数据作为其“食粮”。大数据则像一座藏有无尽知识的图书馆,等待人工智能的发掘与利用。

若将大数据比作图书馆的藏书,人工智能则扮演管理员和研究者的角色,不仅整理书籍(数据),还深入阅读、理解,并通过学习分析,将知识转化为有价值的见解。

两者相辅相成,大数据提供海量信息,人工智能提供处理工具。它们的结合如同图书馆与管理员的配合,高效组织知识,确保信息可靠、真实、稳定,为各种应用提供有力支持。

人工智能特别依赖数据来形成智能,尤其是机器学习。如今,计算能力的提升和大数据技术的发展,使人工智能能够处理海量数据。

(五) 大数据和云计算、物联网、移动互联网、人工智能的关系

大数据和云计算、物联网、移动互联网、人工智能之间均存在密切的关联。

云计算是这些技术的基础,它提供底层服务,包括网络架设、服务器虚拟化、数据存储和计算等。没有云计算的支持,大数据的存储与计算将难以实现。

大数据是这些技术中的一个重要应用,它存在于云计算之上,提供海量数据的分析、整理与汇总功能。没有大数据,云计算就缺少了明确的目标和价值。

物联网、移动互联网以及传统互联网每时每刻都在产生海量的数据,这些数据需要借助大数据技术进行筛选和分析,以便提炼出有价值的信息。

同时,大数据、云计算、物联网、移动互联网等技术的发展都离不开人工智能的参与,以实现更高效、智能的运作。

二、 大数据技术架构

大数据技术是一系列技术的总称,包括数据采集、数据传输、数据存储、数据处理、数据分析、数据挖掘和数据可视化等技术,这些技术共同构成了一个庞大而复杂的技术体系。在实际应用场景中,构建大数据平台时,平台架构的设计及技术选型是至关重要的一步。它们需要根据具体业务需求、数据量、处理性能需求以及预算等因素来综合考虑和选择,这样才能确保大数据平台稳定、高效地运行,从而满足业务需求。

(一) 大数据基础架构

大数据基础架构为四层堆栈式技术架构,包括基础层、管理层、分析层、应用层。

1. 基础层

大数据基础架构的基石是高度自动化的、可扩展的存储和计算平台，包括服务器集群、存储设备(如硬盘、闪存)和网络设备(如交换机、路由器)。

2. 管理层

大数据管理平台整合了结构化和非结构化数据管理，支持实时数据传输、查询和计算。该平台设计时强调并行化和分布式处理，以应对多源数据的深层分析需求。

3. 分析层

分析层集成了统计数据挖掘和机器学习算法，助力企业深度解析数据集，挖掘数据价值。一个灵活、可扩展的分析平台能极大提升数据分析效率。

4. 应用层

大数据的价值在于支持企业决策和为用户提供服务，如智能推荐和广告投放。新兴商业需求推动了大数据应用的发展，同时也加深了企业对大数据价值的认识。

(二) 大数据常用框架技术

大数据框架技术有很多，这里列举一些常用的框架技术。

(1) 文件存储：HDFS(Hadoop Distributed File System, Hadoop 分布式文件系统)、Apache Alluxio、Google Cloud Storage(谷歌云存储)。

(2) 离线计算：Hadoop MapReduce(Hadoop 映射归约)、Spark(Apache Spark)。

(3) 流式、实时计算：Apache Storm、Spark Streaming(Apache Spark Streaming, 火花流处理)、Apache Flink。

(4) K-V、NoSQL 和关系数据库：HBase(Apache HBase)、Redis、MongoDB、MySQL 关系型数据库管理系统)。

(5) 资源管理：YARN(Yet Another Resource Negotiator, 另一种资源协调者)、Mesos(Apache Mesos)。

(6) 日志收集：Apache Flume、Logstash、Elastic Stack。

(7) 消息系统：Apache Kafka、RabbitMQ。

(8) 查询分析：Hive、Trino(原 PrestoSQL)、Apache Impala。

(9) 分布式协调服务：Apache ZooKeeper。

(10) 集群管理与监控：Apache Ambari、Cloudera Manager。

(11) 数据挖掘、机器学习：Apache Mahout、Spark MLlib(Apache Spark MLlib, 火花机器学习库)。

(12) 数据同步：Apache Sqoop。

(13) 任务调度：Apache Oozie。

三、 大数据的基本处理流程

大数据的基本处理流程主要包括数据采集与预处理、数据存储与管理、数据处理与分析、数据可视化、数据安全与隐私保护等，如表 1-2 所示。数据可视化有时被视为数据分析的一部分，即可视化分析，因此，数据可视化也可归入数据处理与分析。

表 1-2 大数据基本处理流程及其描述

处理流程	描述
数据采集与预处理	利用 ETL(Extraction-Transformation-Loading, 抽取—转换—加载工具), 我们能够从异构数据源(如关系数据库)中高效地抽取数据, 随后进行精细的清洗、转换和集成处理, 最终将数据加载至数据仓库或数据集, 从而为联机分析处理(Online Analytical Processing, OLAP)和数据挖掘任务奠定坚实的基础。此外, 借助日志采集工具(如 Flume 等), 我们还能实时捕获数据流, 并将其作为流计算系统的输入, 实现实时、高效的数据处理与分析
数据存储与管理	通过整合分布式文件系统(如 HDFS)、高性能数据仓库(如 Hive)、关系型数据库(如 MySQL)、NoSQL 数据库(如 Cassandra、HBase)以及云数据库服务, 实现对结构化、半结构化和非结构化数据的全面、高效存储与管理。这些技术的融合确保了数据的灵活性、可扩展性和安全性, 满足了不同业务场景下的数据存储需求
数据处理与分析	借助分布式并行编程模型(如 MapReduce)以及高效计算框架(如 Apache Spark、Apache Flink), 结合前沿的机器学习和数据挖掘算法, 我们能够实现对海量数据精准处理与深入分析, 从而高效提取出有价值的信息
数据可视化	将分析结果以图表、图像和仪表板等直观形式进行可视化展现, 助力用户深入理解数据、精准分析趋势, 进而基于数据作出更加明智的决策
数据安全与隐私保护	在挖掘大数据的商业与学术价值时, 我们需要高度重视隐私数据保护, 通过加密技术、数据脱敏及严格的访问控制策略, 筑牢数据安全防线, 全力保障个人隐私和数据安全

每种大数据技术都有其独特的适用场景, 因此, 企业在选择大数据解决方案时, 应首先明确其应用场景, 从而精准匹配最适宜的大数据计算模式, 进而挑选出与之相匹配的大数据计算产品。



描述物联网技术在智能家居中的具体应用, 如智能照明、智能安防、智能家电等。



实训 使用百度指数体验大数据

实训目标

百度指数是基于百度庞大的用户行为数据而开发的数据分析工具，它在互联网及数据时代扮演着统计分析角色，自推出以来便成为企业制定营销策略的重要参考。

该平台能够清晰地揭示特定关键词在百度上的搜索量、波动趋势以及相关新闻舆论的动态变化；同时，它还能详细展示关注这些关键词的用户特征、地域分布以及他们的其他相关搜索行为。通过这些信息，百度指数可以有效地帮助用户提升数字营销的效果。

本实训旨在通过百度指数工具，对关键词“专升本”进行详细的搜索趋势、资讯指数、需求图谱、人群画像等多维度分析，以了解当前社会对“专升本”的关注度、趋势变化、相关需求以及关注人群的特征。

实训过程

一、趋势研究

（一）搜索指数

在浏览器中输入百度指数网址(<http://index.baidu.com>)，进入百度指数首页，如图 1-5 所示。

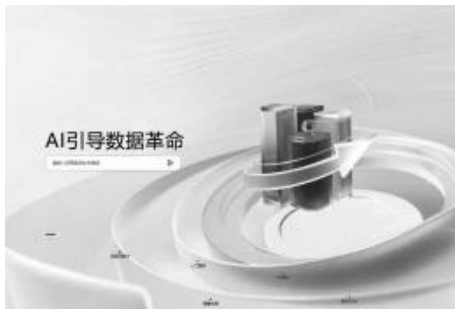


图 1-5 百度指数首页

在搜索框中输入关键词“专升本”，单击探索按钮，搜索指数结果将展示与“专升本”相关的数据，如图 1-6 所示。



图 1-6 “专升本”的搜索指数图(2024 年 5 月 6 日至 2024 年 6 月 4 日)

搜索指数默认的时间范围为近 30 天，但可以根据需要在页面上方的日期选择器或下方的时间轴上调整时间范围。图 1-6 显示的是近 30 天“专升本”的搜索指数。

总体来看，2024 年 5 月 6 日至 2024 年 6 月 4 日，PC 端与移动端的“专升本”搜索指数呈现一种动态波动趋势，其日平均搜索量稳定在 16 819，特别是在 6 月初，随着学生面临的就业市场压力与学历提升需求的增加，关于“专升本”的讨论度和搜索量会相应上升。此外，随着国家政策支持和教育改革的深入，专升本教育的普及度不断提高。由图 1-6 提供的相关数据可以看到，近一个月的“专升本”搜索指数较去年同比上升 212%，较上一个月环比上升 18%。

(二) 资讯指数

在“搜索指数”图的下方是“专升本”资讯指数图，如图 1-7 所示。

近 30 天内出现 7 次“专升本”的内容头条，依次以 A、B、C、D、E、F 和 G 进行标注，单击“F”可出现具体内容信息。如“专升本人绝不会说的 10 个小秘密”等。

根据“资讯指数概览”提供的相关数据可知，专升本的资讯指数日均值为 189 117，较去年同比下降 17%，而较上月环比下降 14%。

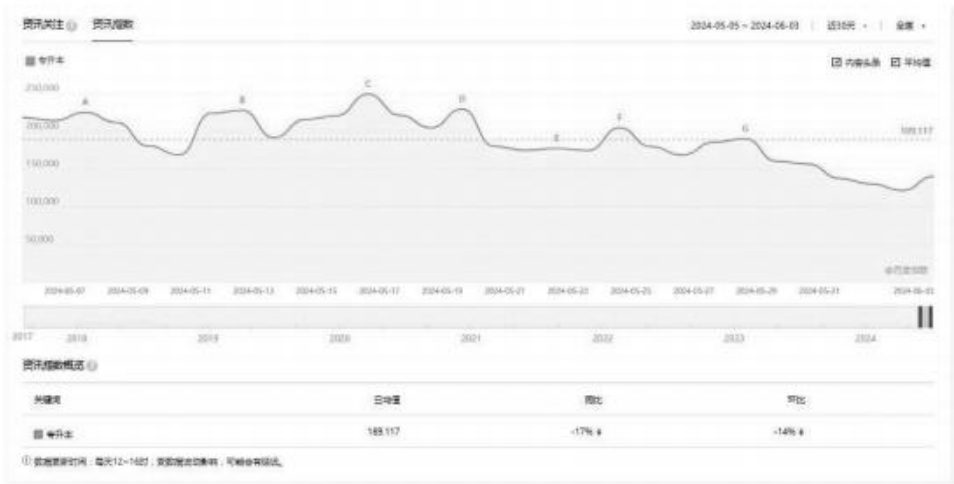
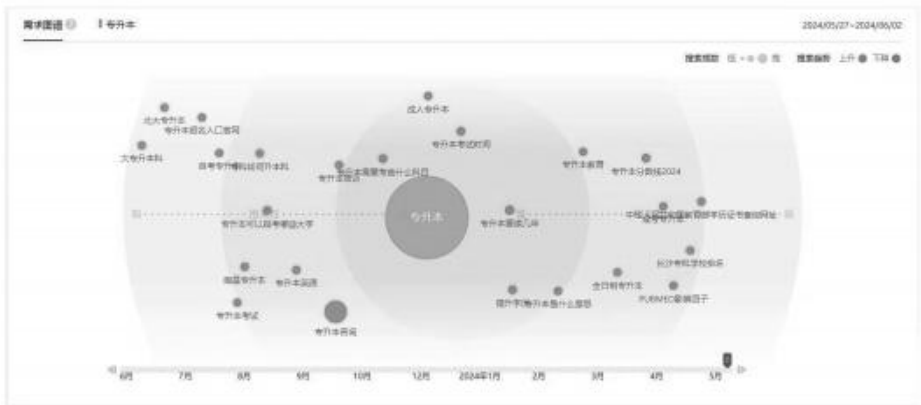


图 1-7 “专升本”的资讯指数图

二、需求图谱

(一) 需求分布

时间段选择不同，需求图谱展示出相应的结果也不相同，此处以 2024 年 5 月 27 日至 2024 年 6 月 2 日的需求图谱为例，如图 1-8 所示。



扫码查看 1-8 彩图

图 1-8 “专升本”的需求图谱

由图 1-8 可知，在这一时间段与“专升本”相关性最强的词分别为“专升本需要考些什么科目”“专升本要读几年”“专升本考试时间”“成人专升本”“提升学历”“专升本培训”等。

相关性最强的词中自身搜索指数最高的是“专升本咨询”，并且搜索呈上升趋势，“专升本培训”的搜索呈下降趋势。“大专升本科”与“中华人民共和国教育部学历证书查询网址”搜索相关性最弱，在该时间段内搜索呈下降趋势。

（二）相关词

2024 年 5 月 27 日至 2024 年 6 月 2 日，相关词热度如图 1-9 所示。



图 1-9 “专升本”的相关词热度

相关词中“专升本咨询”的搜索热度最高，其次是“成人本科”“专升本要读几年”“专升本第一学历是本科还是专科”等。搜索率上升幅度最大的相关词为“专升本咨询”，其次是“专科如何升本科”“专升本培训”“大专升本科”等。搜索率下降的相关词为“专升本成绩查询”“专升本教育”“成人高考报考”“专升本分数线 2024”“南昌专升本”“专升本报名入口官网”。

三、人群画像

（一）地域分布

2024 年 5 月 5 日至 2024 年 6 月 4 日，关注“专升本”的用户主要来自广东、四川、江苏、山东、浙江等省份。我国华东地区对“专升本”的关注度要高于我国西北部地区，如图 1-10 所示。根据需要可调整时间段，最早可追溯至 2013 年 7 月 1 日。

（二）人群属性

百度指数给出关注“专升本”的人群属性，如图 1-11 所示，数据选取时间段为 2024 年 5 月 1 日至 2024 年 5 月 31 日。



图 1-10 关注“专升本”的地域分布

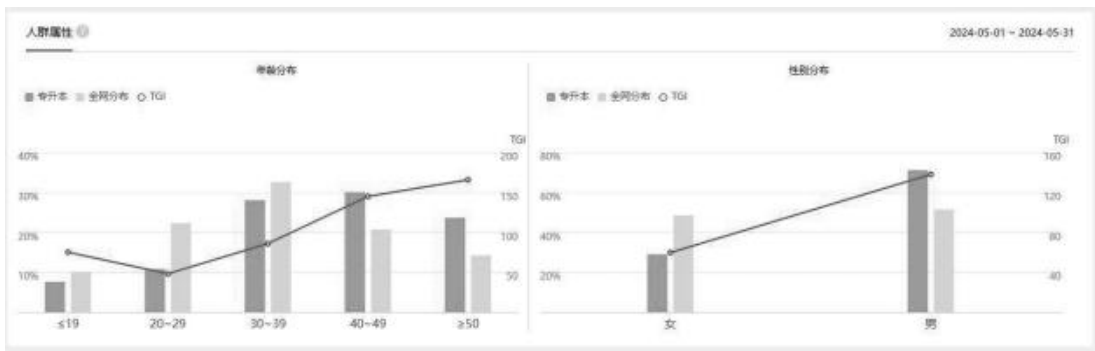


图 1-11 关注“专升本”的人群属性

在此期间关注“专升本”的人群中年龄段在 40~49 岁的人最多，年龄在 19 岁以下的人最少，其中女性占 59.67%。TGI 指数表明不同特征用户关注问题的差异情况，其中 TGI 指数等于 100 表示平均水平，高于 100 代表该类用户对某类问题的关注程度高于整体水平。

从图 1-11 中可以看出，年龄超过 40 岁的人群对“专升本”的关注度超过整体水平，50 岁以上人群的 TGI 指数高达 165.79，男性的 TGI 指数高达 138.01。

（三）兴趣分布

百度指数还给出了关注“专升本”的人群兴趣分布结果，如图 1-12 所示，数据选取时间段为 2024 年 5 月 1 日至 2024 年 5 月 31 日。

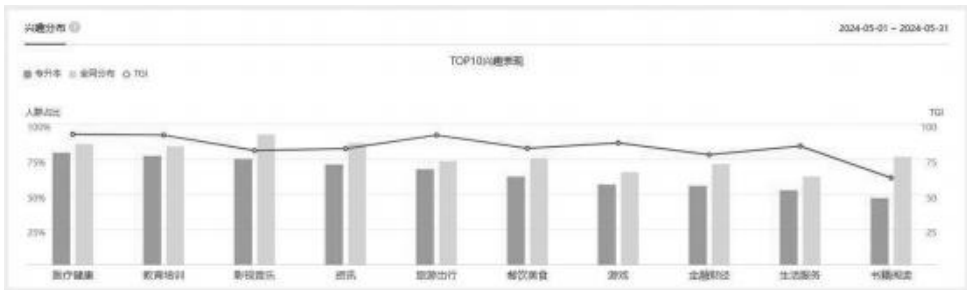


图 1-12 关注“专升本”人群的兴趣分布

从“医疗健康”“教育培训”“金融财经”等维度对关注“专升本”的人群进行分类，以关注“金融财经”的人群为例，该类人群中有 55.78% 的人同时关注了“专升本”，全网分布 71.53%，TGI 指数为 77.98。

请根据实训情况填写自评表，见表 1-3 和表 1-4。

表 1-3 学生技能自评表

序号	技能	描述	自评等级
1	数据分析技能	能够熟练运用数据分析工具(如百度指数)进行关键词搜索指数、资讯指数等数据的查询与分析,从而洞察市场趋势和用户需求	优秀/良好/一般/较差
2	搜索引擎应用能力	熟练掌握搜索引擎的使用方法,能够高效、准确地从海量信息中检索出所需数据和信息,支持决策制定和问题研究	优秀/良好/一般/较差
3	图表解读能力	能够读懂并解释图表中的信息,如搜索指数趋势图、需求图谱等,提取关键数据点,为分析提供有力支持	优秀/良好/一般/较差

注:自评等级说明

- ①优秀:在相关评价项目中表现出色,达到预期目标并超出预期。
- ②良好:在相关评价项目中表现良好,达到预期目标。
- ③一般:在相关评价项目中表现一般,基本达到预期目标但存在不足。
- ④较差:在相关评价项目中表现不佳,未达到预期目标。

表 1-4 学生素养自评表

序号	素养	描述	自评等级
1	批判性思维	具备独立思考的能力,能够客观地分析数据和信息,不盲从,对问题有自己的见解和判断	优秀/良好/一般/较差
2	信息素养	具备良好的信息搜集、整理、分析和应用能力,能够有效地利用信息资源进行学习和研究	优秀/良好/一般/较差
3	团队协作能力	在实训过程中,能够与他人有效沟通,共同完成任务,展现出良好的团队合作精神和协作能力	优秀/良好/一般/较差



一、单项选择题

1. 下列属于数据存储单位的是 ()。
A. XB B. QB C. TB D. CB
2. 关于数字经济的基本特征， 以下描述最为准确的一项是 ()。
A. 数字经济主要依赖于传统的生产方式和商业模式
B. 数字经济通过减少中间环节和交易成本， 实现了高效直联性
C. 数字经济中， 摩尔定律揭示了网络价值的本质
D. 数字经济中的边际收益递减性显著， 导致规模经济效应减弱
3. 下列不是大数据在互联网领域的主要应用的是 ()。
A. 个性化推荐 B. 精准营销 C. 广告投放 D. 物流跟踪
4. 大数据思维的方式不包括 ()。
A. 总体性思维 B. 容错性思维
C. 精确性思维 D. 智能化思维
5. 物联网的体系架构中， 负责识别和采集各种物体的信息的一层是 ()。
A. 感知层 B. 网络层 C. 应用层 D. 数据层
6. 在大数据处理流程中， 以下步骤中主要关注从异构数据源中抽取数据， 并进行清洗、 转换和集成处理的是 ()。
A. 数据采集与预处理 B. 数据存储与管理
C. 数据处理与分析 D. 数据可视化
7. 在数据存储与管理中， 以下主要用于存储大规模非结构化数据的技术是 ()。
A. HDFS B. Hive C. MySQL D. Cassandra

二、多项选择题

1. 数据的特征包括 ()。
A. 差异性 B. 相似性 C. 规律性 D. 随机性
2. 以下属于按性质分类的数据有 ()。
A. 定量数据 B. 文本数据 C. 定性数据 D. 实时数据
3. 电信领域大数据的应用主要包括 ()。
A. 网络管理与优化 B. 市场与精准营销
C. 客户关系管理 D. 企业运营管理
E. 社交媒体分析

4. 大数据思维在应用方面具有显著优势的领域有 ()。
- A. 科学研究 B. 市场营销 C. 医疗健康 D. 制造业
5. 云计算根据服务模式可以划分的类型有 ()。
- A. 基础设施即服务 (IaaS) B. 平台即服务 (PaaS)
- C. 软件即服务 (SaaS) D. 数据即服务 (DaaS)
6. 在大数据应用中, 通常用于文件存储的技术有 ()。
- A. Hadoop HDFS B. Apache Storm
- C. MySQL D. Google Cloud Storage

三、判断题

1. 数据存储单位从 byte 到 BB 的换算关系都是基于 2 的幂次方。()
2. 大数据时代, 结构化数据在总数据量中占据主要比重。()
3. 电商平台通过大数据能够精准预测用户的购买需求。()
4. 大数据思维仅仅关注数据的数量, 而不关注数据的质量。()
5. 在大数据时代, 人们不再需要依赖样本数据进行分析和预测。()
6. 在设计大数据管理平台时, 只需要考虑数据的存储和管理, 不需要考虑数据的计算。()