



《大数据导论》

教材官网: <http://dblab.xmu.edu.cn/post/bigdata-introduction/>

温馨提示: 编辑幻灯片母版, 可以修改每页PPT的厦大校徽和底部文字

第2章 大数据与其他新兴技术的关系

(PPT版本号: 2020年秋季学期)



扫一扫访问教材官网

林子雨 博士/副教授

厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn ▶▶

主页: <http://dblab.xmu.edu.cn/post/linziyu>





课程教材

- 林子雨 编著 《大数据导论》
 - 人民邮电出版社，2020年9月第1版
 - ISBN:978-7-115-54446-9 定价：49.80元
- 教材官网：<http://dbllab.xmu.edu.cn/post/bigdata-introduction/>



扫一扫访问教材官网



提纲

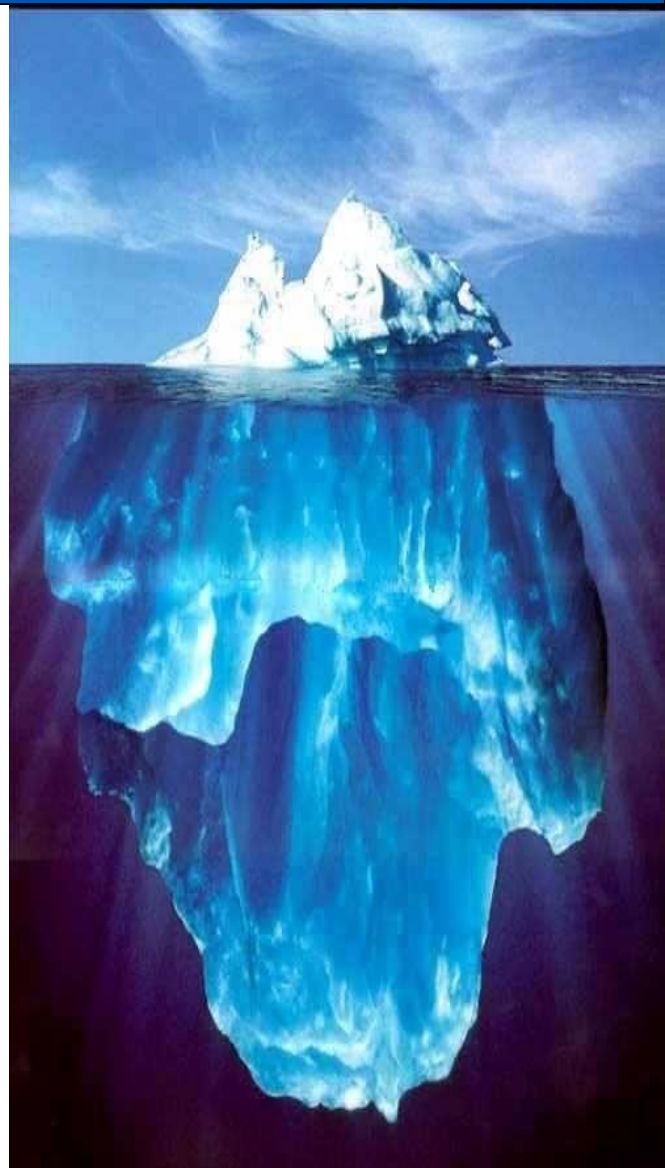
- 2.1 云计算
- 2.2 物联网
- 2.3 大数据与云计算、物联网的关系
- 2.4 大数据与人工智能
- 2.5 大数据与区块链



高校大数据课程

公共服务平台

百度搜索厦门大学数据库实验室网站访问平台





2.1 云计算

2.1.1 云计算概念

2.1.2 云计算服务模式 and 类型

2.1.3 云计算数据中心

2.1.4 云计算的应用

2.1.5 云计算产业



2.1.1 云计算概念



云计算示意图

云计算概念

- 通过网络、以服务的方式，为千家万户提供非常廉价的IT资源

云计算特点

- 超大规模计算、虚拟化、高可靠性和安全性、通用性、动态扩展性、按需服务、降低成本



2.1.1 云计算概念

水资源获取方式





2.1.1 云计算概念

水资源获取方式——挖井取水



- (1) 初期成本高，周期长
- (2) 后期需要自己维护
- (3) 供水量有限



2.1.1 云计算概念

水资源获取方式——自来水



- (1) 初期零成本，瞬时可获得
- (2) 后期免维护，使用成本低
- (3) 在供水量方面“予取予求”



2.1.1 云计算概念

传统IT资源获取方式：自建机房



- (1) 初期成本高，周期长
- (2) 后期需要自己维护，使用成本高
- (3) IT资源供应量有限



2.1.1 云计算概念

云计算

企业不需要自建IT基础设施，可以租用云端资源



无需自建

- (1) 初期零成本，瞬时可获得
- (2) 后期免维护，使用成本低
- (3) 在供应IT资源量方面“予取予求”

企业用户

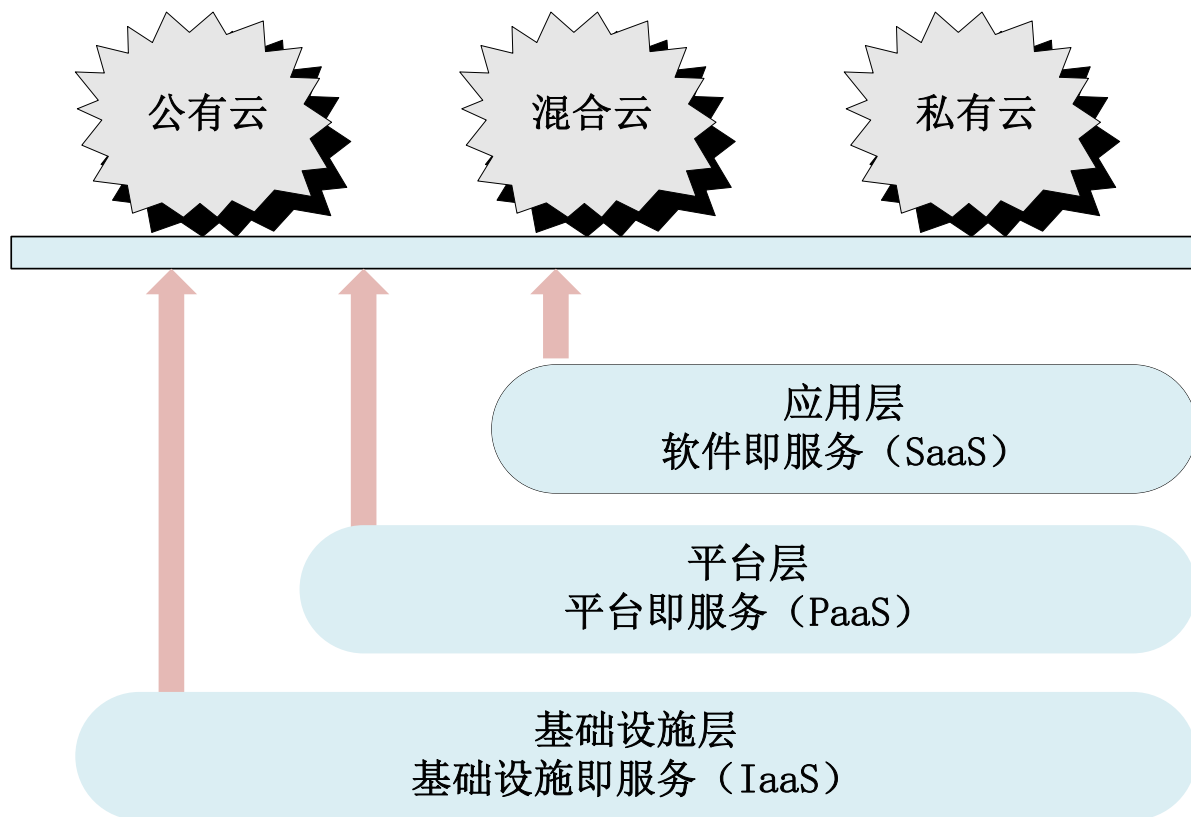


租用云端资源





2.1.2 云计算服务模式 and 类型





2.1.2 云计算服务模式 and 类型

SaaS

从一个集中的系统部署软件，使之在一台本地计算机上(或从云中远程地)运行的一个模型。由于是计量服务，SaaS 允许出租一个应用程序，并计时收费

PaaS

类似于 IaaS，但是它包括操作系统和围绕特定应用的必需的服务

IaaS

将基础设施(计算资源和存储)作为服务出租

Application

Platform

Infrastructure

Visualization

Server

Storage

Server

Storage

SaaS

Software as a Service

Google Apps, Microsoft “Software+Services”

PaaS

Platform as a Service

IBM IT factory, Google App Engine, Force.com

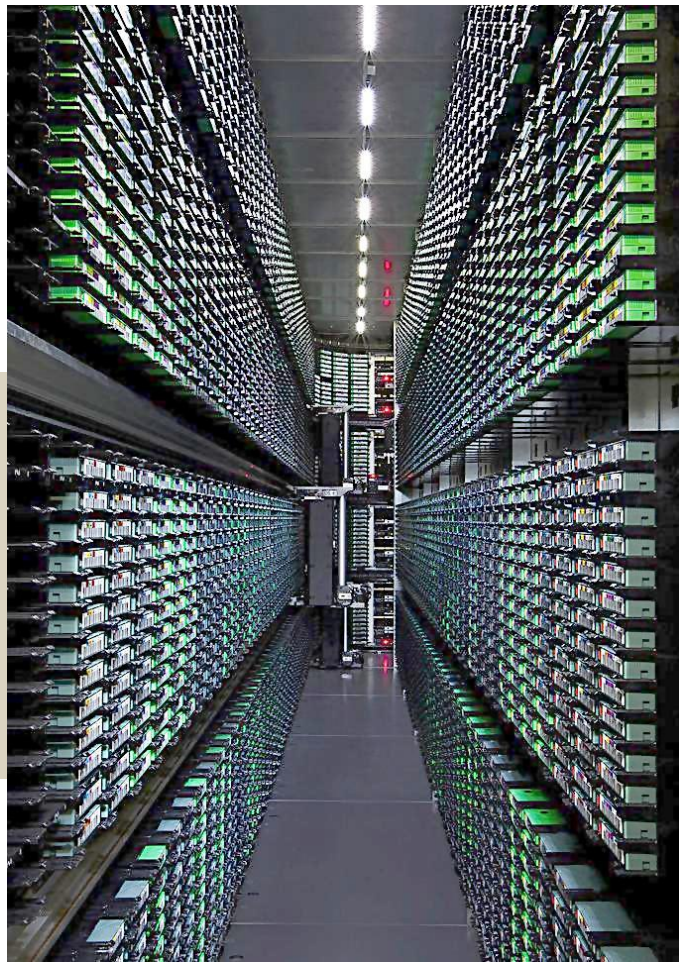
IaaS

Infrastructure as a Service

Amazon EC2, IBM Blue Cloud, Sun Grid



2.1.3 云计算数据中心



- 数据中心是云计算的温床
- 云计算推动数据中心向虚拟化和云架构的转型，不断提高IT基础架构的灵活性，以降低IT、能源和空间成本，从而让客户能够快速提高业务敏捷性



“那些正在兴建最大规模数据中心的公司一般都在云计算方面拥有巨大野心”

-- 《数据中心知识》杂志主编 Rich Miller



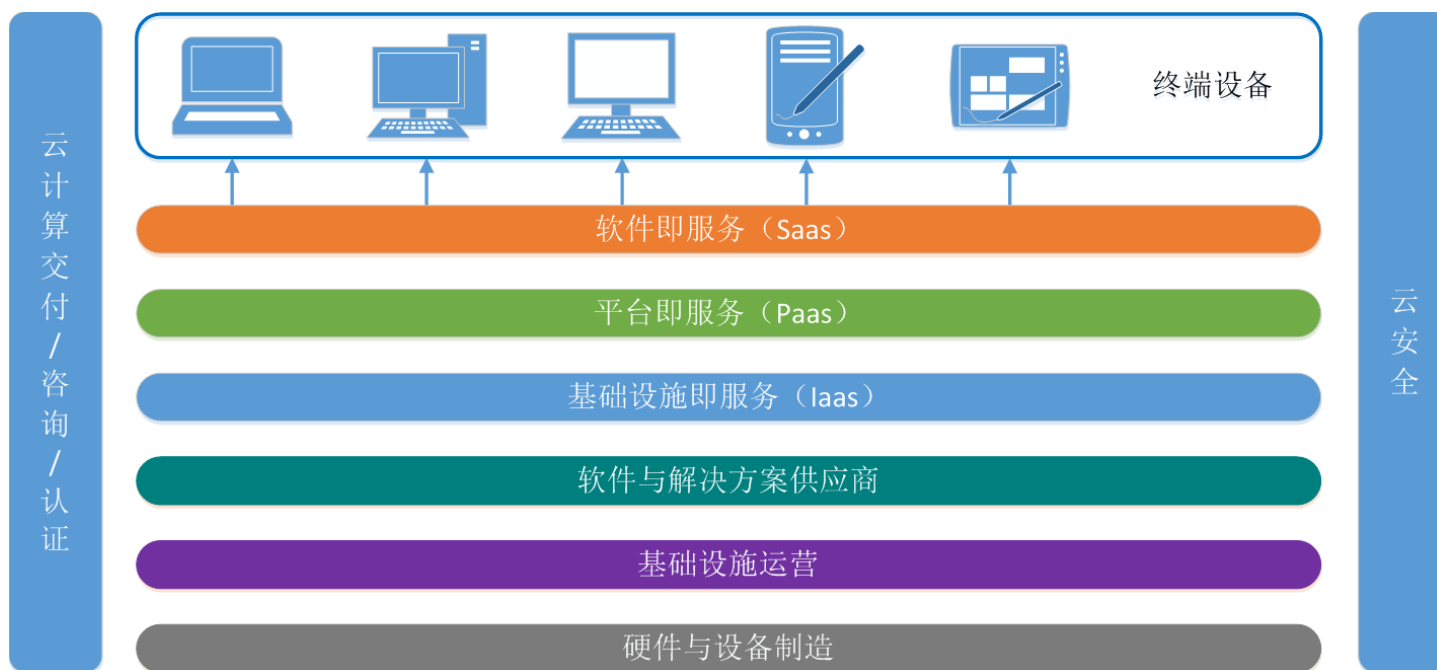
2.1.4 云计算的应用

- 政务云上可以部署公共安全管理、容灾备份、城市管理、应急管理、智能交通、社会保障等应用，通过集约化建设、管理和运行，可以实现信息资源整合和政务资源共享，推动政务管理创新，加快向服务型政府转型
- 教育云可以有效整合幼儿教育、中小学教育、高等教育以及继续教育等优质教育资源，逐步实现教育信息共享、教育资源共享及教育资源深度挖掘等目标
- 中小企业云能够让企业以低廉的成本建立财务、供应链、客户关系等管理应用系统，大大降低企业信息化门槛，迅速提升企业信息化水平，增强企业市场竞争力
- 医疗云可以推动医院与医院、医院与社区、医院与急救中心、医院与家庭之间的服务共享，并形成一套全新的医疗健康服务系统，从而有效地提高医疗保健的质量



2.1.5 云计算产业

- 云计算产业作为战略性新兴产业，近些年得到了迅速发展，形成了成熟的产业链结构，产业涵盖硬件与设备制造、基础设施运营、软件与解决方案供应商、基础设施即服务（IaaS）、平台即服务（PaaS）、软件即服务（SaaS）、终端设备、云安全、云计算交付/咨询/认证等环节





2.2 物联网

2.2.1 物联网的概念

2.2.2 物联网关键技术

2.2.3 物联网的应用

2.2.4 物联网产业



2.2.1 物联网的概念

- 物联网是物物相连的互联网，是互联网的延伸，它利用局部网络或互联网等通信技术把传感器、控制器、机器、人员和物等通过新的方式联在一起，形成人与物、物与物相联，实现信息化和远程管理控制

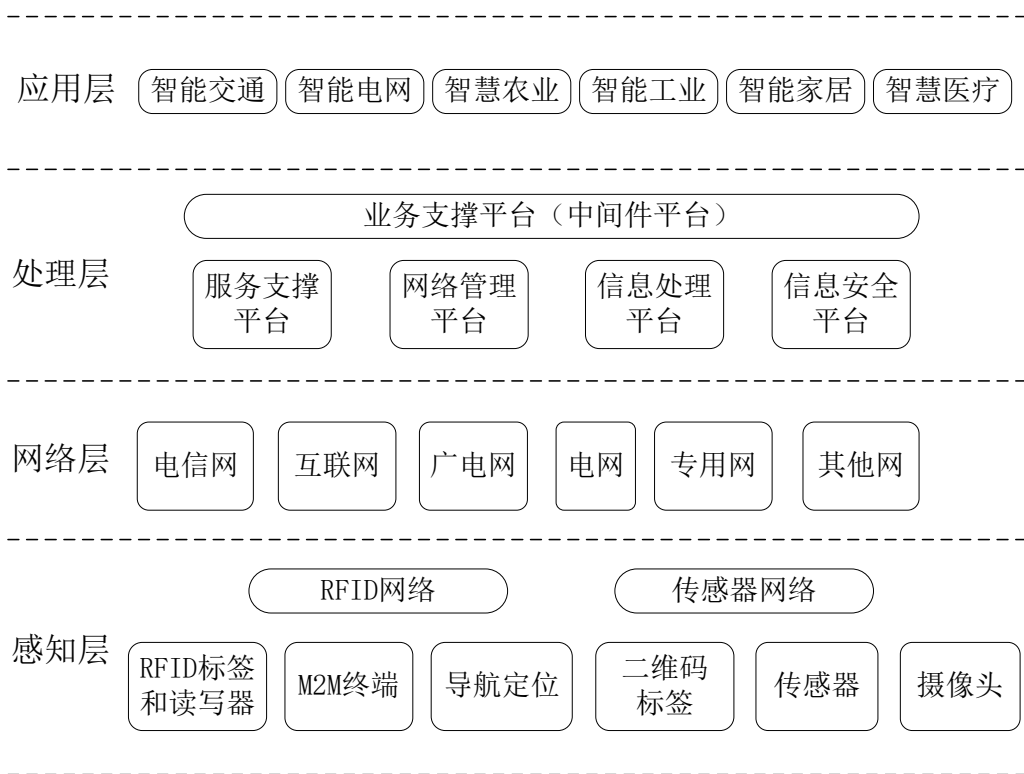


图 物联网体系架构



2.2.2 物联网关键技术

- 物联网中的关键技术包括识别和感知技术（二维码、RFID、传感器等）、网络与通信技术、数据挖掘与融合技术等

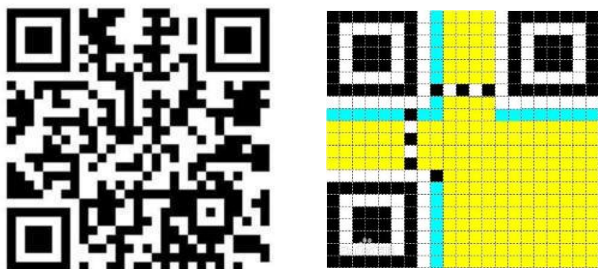


图 矩阵式二维码



图 采用RFID芯片的公交卡



(a)温湿度传感器



(b)压力传感器



(c)烟雾传感器

图 不同类型的传感器



2.2.3 物联网的应用

- 物联网已经广泛应用于智能交通、智慧医疗、智能家居、环保监测、智能安防、智能物流、智能电网、智慧农业、智能工业等领域，对国民经济与社会发展起到了重要的推动作用





2.2.4 物联网产业

- 完整的物联网产业链主要包括核心感应器件提供商、感知层末端设备提供商、网络提供商、软件与行业解决方案提供商、系统集成商、运营及服务提供商等六大环节

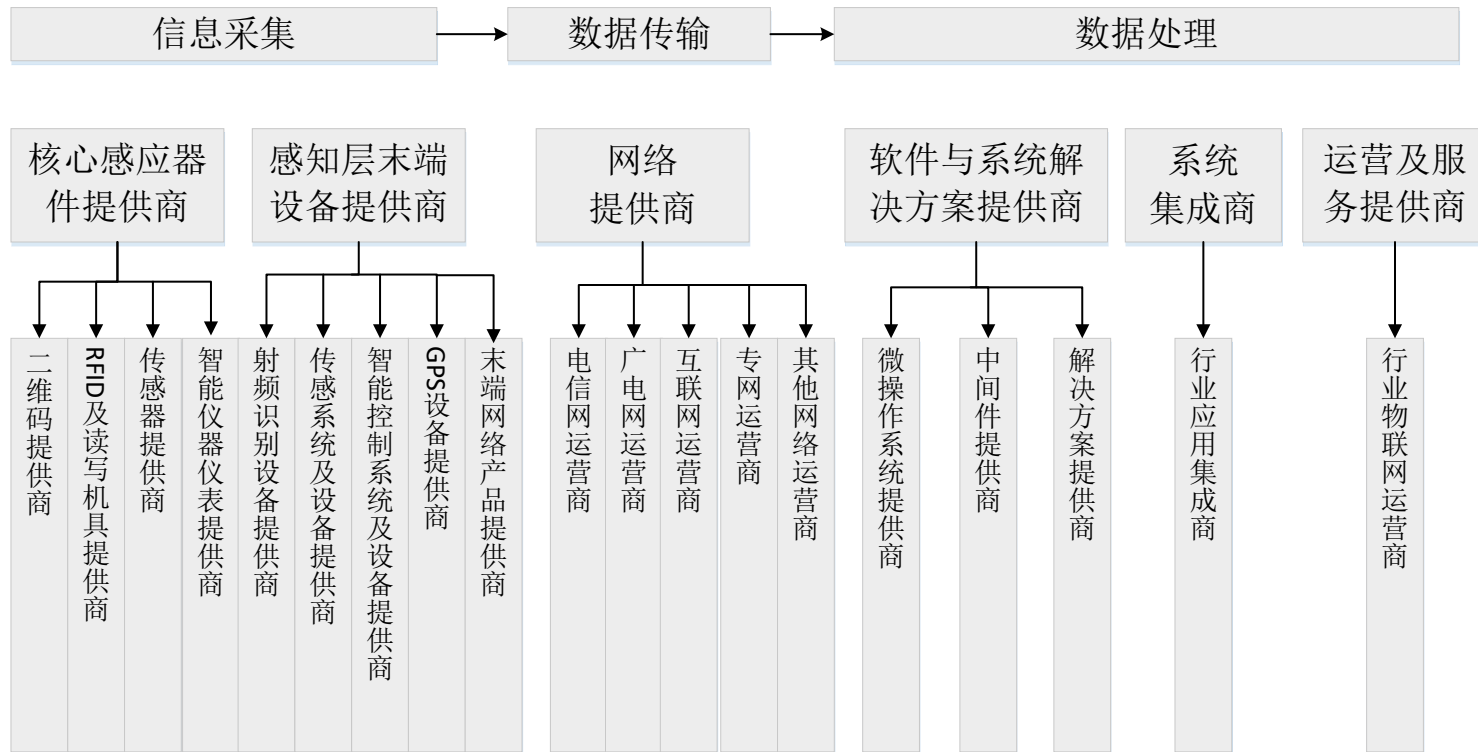


图 物联网产业链



2.3 大数据与云计算、物联网的关系

- 云计算、大数据和物联网代表了IT领域最新的技术发展趋势，三者既有区别又有联系

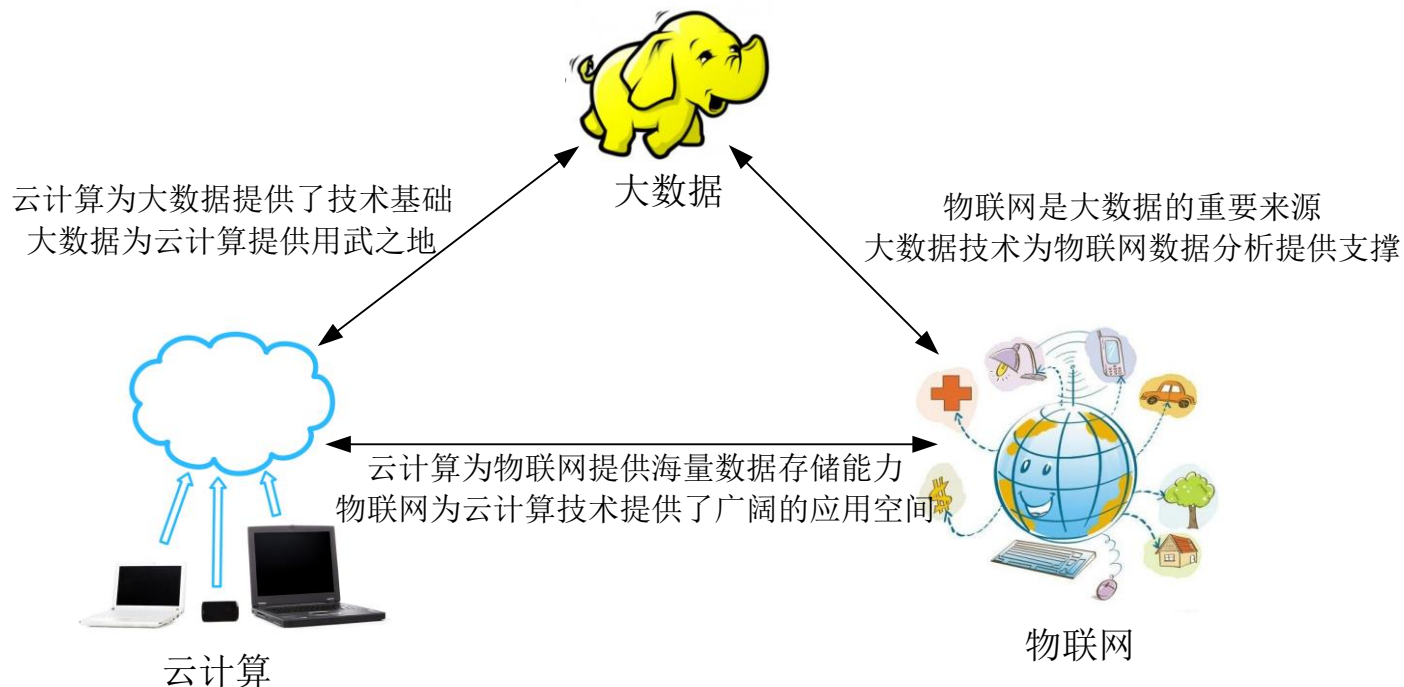


图 大数据、云计算和物联网之间的关系



2.4 大数据与人工智能

2.4.1 人工智能概念

2.4.2 人工智能关键技术

2.4.3 人工智能的应用

2.4.4 人工智能产业

2.4.5 大数据与人工智能的关系

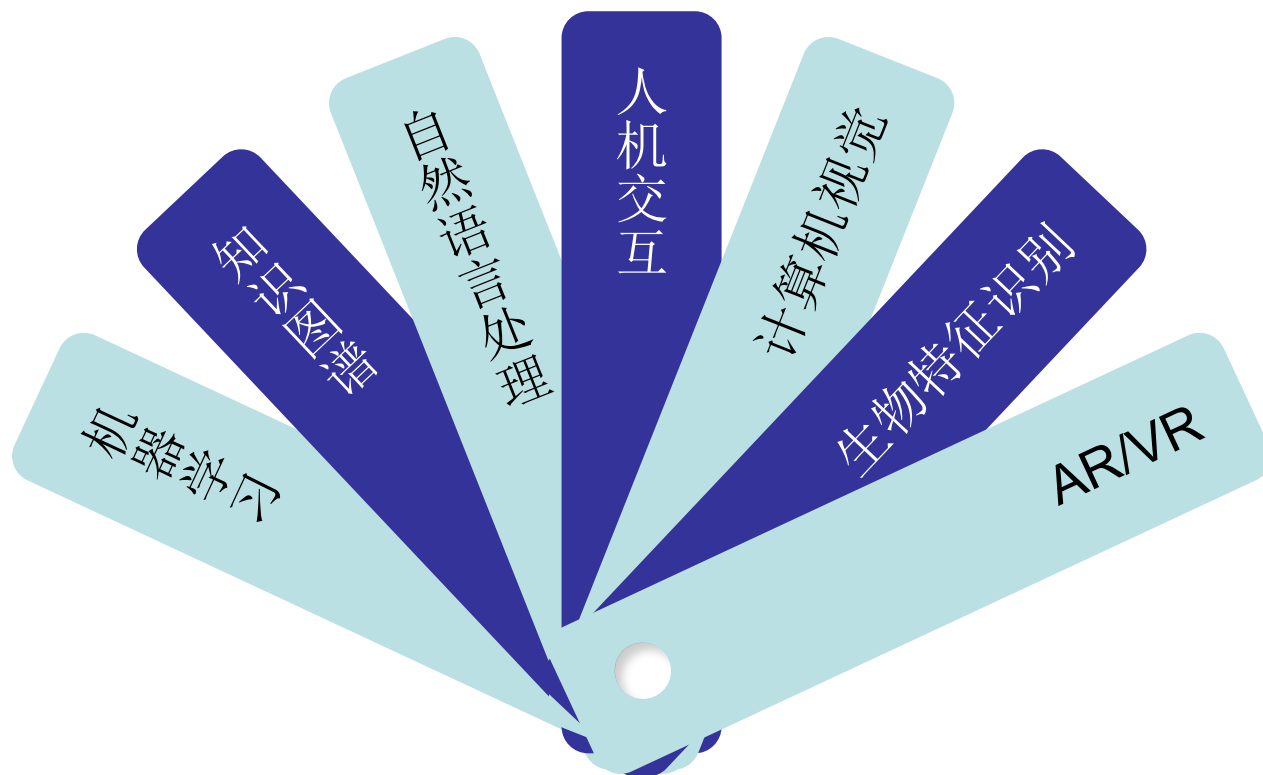


2.4.1 人工智能概念

- 人工智能（**Artificial Intelligence**），英文缩写为**AI**，是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。
- 人工智能是计算机科学的一个分支，它企图了解智能的实质，并生产出一种新的能以与人类智能相似的方式做出反应的智能机器，该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。



2.4.2 人工智能关键技术





2.4.2 人工智能关键技术

1. 机器学习

强调三个关键词：算法、经验、性能，其处理过程如图所示。在数据的基础上，通过算法构建出模型并对模型进行评估。评估的性能如果达到要求，就用该模型来测试其他的数据；如果达不到要求，就要调整算法来重新建立模型，再次进行评估。如此循环往复，最终获得满意的模型来处理其他数据。

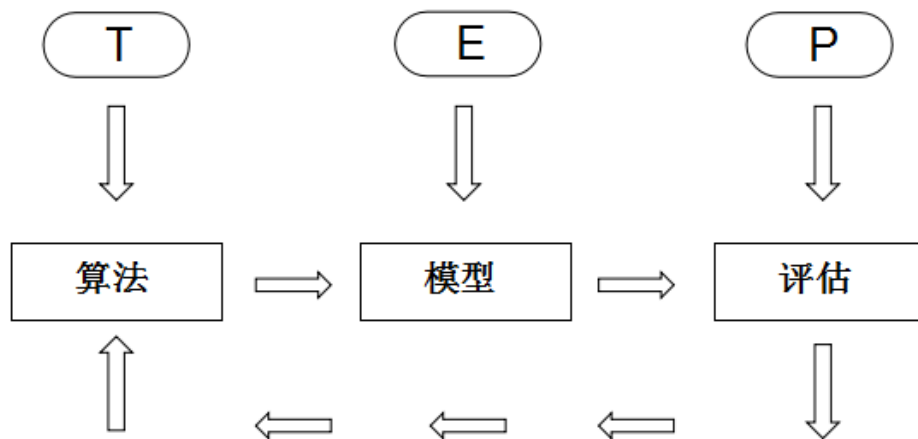


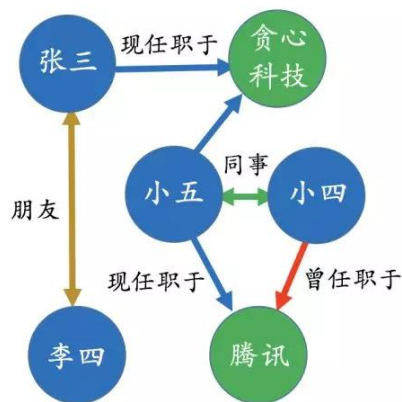
图2-13 机器学习处理过程



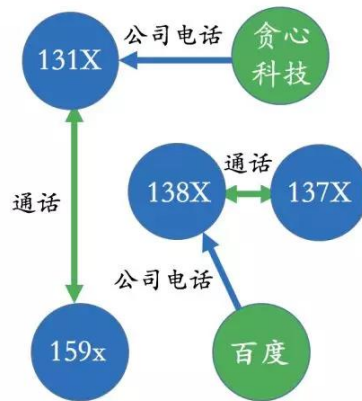
2.4.2 人工智能关键技术

2.知识图谱

知识图谱（Knowledge Graph）又称为科学知识图谱，在图书情报界称为知识域可视化或知识领域映射地图，是显示知识发展进程与结构关系的一系列各种不同的图形，用可视化技术描述知识资源及其载体，挖掘、分析、构建、绘制和显示知识及它们之间的相互联系。



案例：社交网络



案例：风控知识图谱



2.4.2 人工智能关键技术

3. 自然语言处理

- 自然语言处理是计算机科学领域与人工智能领域中的一个重要方向。它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。
- 自然语言处理的应用包罗万象，例如：机器翻译、手写体和印刷体字符识别、语音识别、信息检索、信息抽取与过滤、文本分类与聚类、舆情分析和观点挖掘等



2.4.2 人工智能关键技术

4.人机交互

- 人机交互是一门研究系统与用户之间的交互关系的学科。系统可以是各种各样的机器，也可以是计算机化的系统和软件。
- 人机交互界面通常是指用户可见的部分。用户通过人机交互界面与系统交流，并进行操作。
- 人机交互是与认知心理学、人机工程学、多媒体技术、虚拟现实技术等密切相关的综合学科。

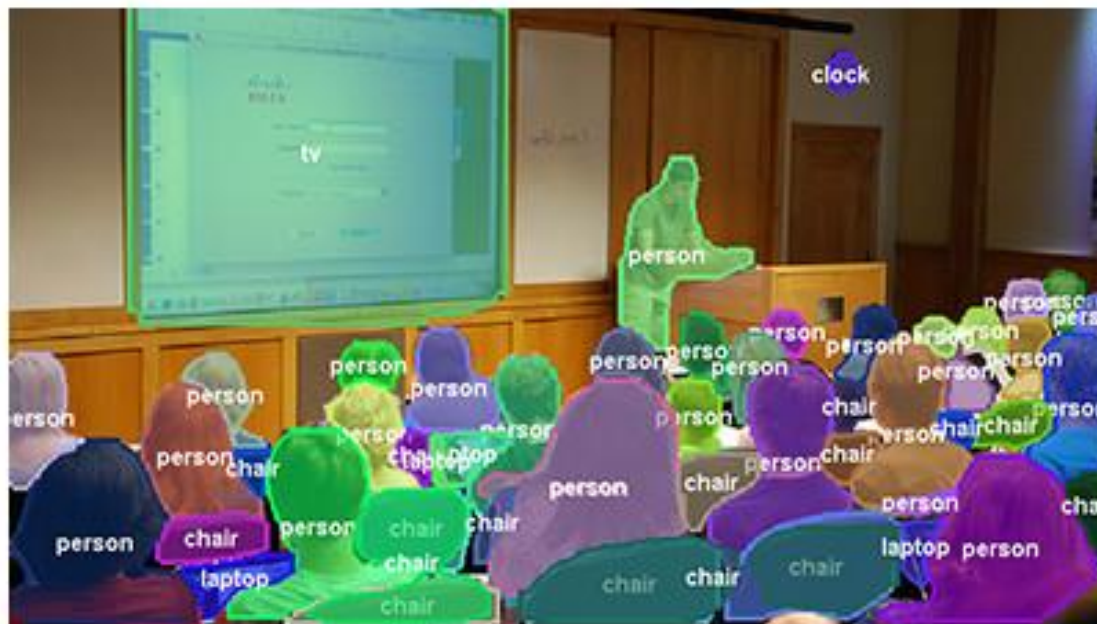




2.4.2 人工智能关键技术

5. 计算机视觉

计算机视觉是一门研究如何使机器“看”的科学，更进一步地说，是指用摄影机和电脑代替人眼对目标进行识别、跟踪和测量的机器视觉，并进一步做图形处理，成为更适合人眼观察或传送给仪器检测的图像

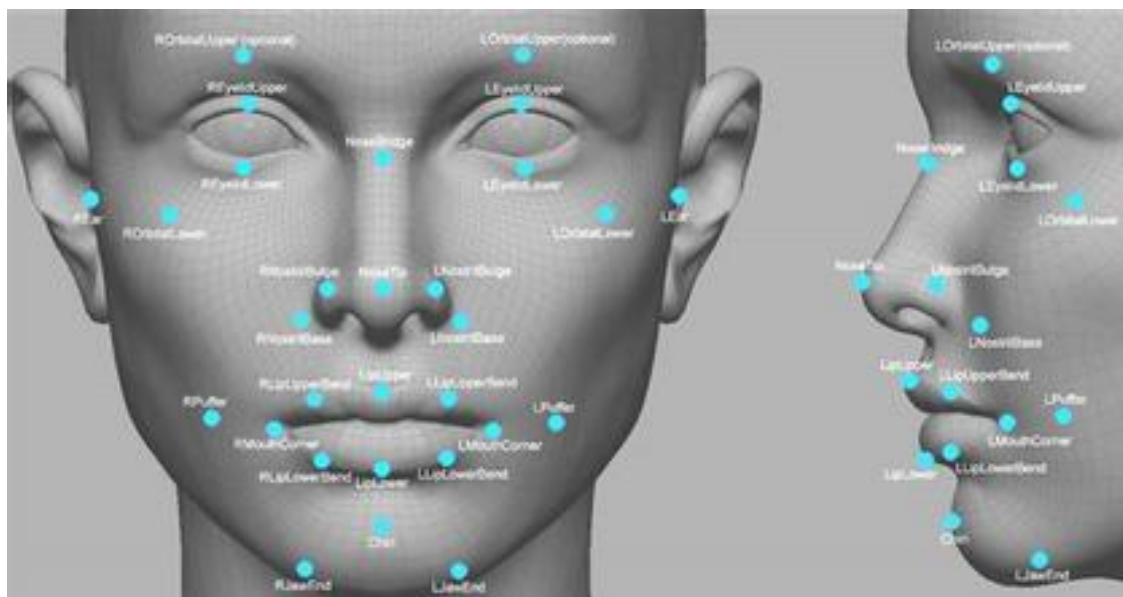




2.4.2 人工智能关键技术

6.生物特征识别

生物特征识别技术涉及的内容十分广泛，包括指纹、掌纹、人脸、虹膜、指静脉、声纹、步态等多种生物特征，其识别过程涉及到图像处理、计算机视觉、语音识别、机器学习等多项技术





2.4.2 人工智能关键技术

7.VR/AR

虚拟现实（VR）/增强现实（AR）是以计算机为核心的新型视听技术。结合相关科学技术，在一定范围内生成与真实环境在视觉、听觉、触感等方面高度近似的数字化环境。



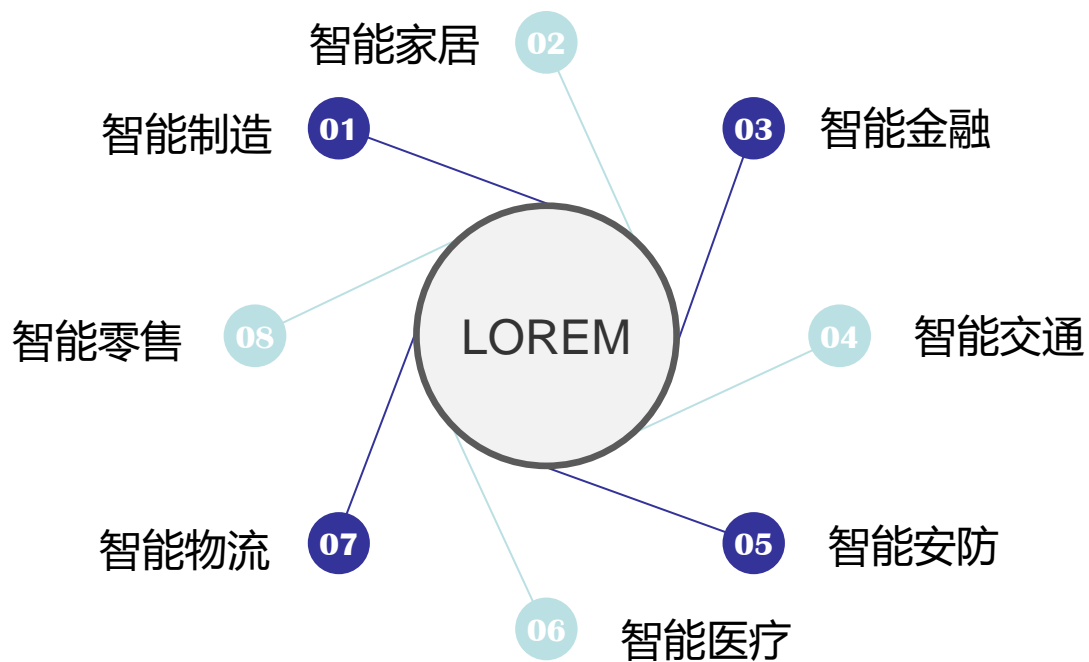
采用虚拟现实技术的虚拟弓箭



利用Tilt Brush在虚拟三维空间里绘画



2.4.3 人工智能的应用





2.4.3 人工智能的应用

1. 智能制造

智能制造（Intelligent Manufacturing, IM）是一种由智能机器和人类专家共同组成的人机一体化智能系统，它在制造过程中能进行智能活动，诸如分析、推理、判断、构思和决策等





2.4.3 人工智能的应用

2.智能家居

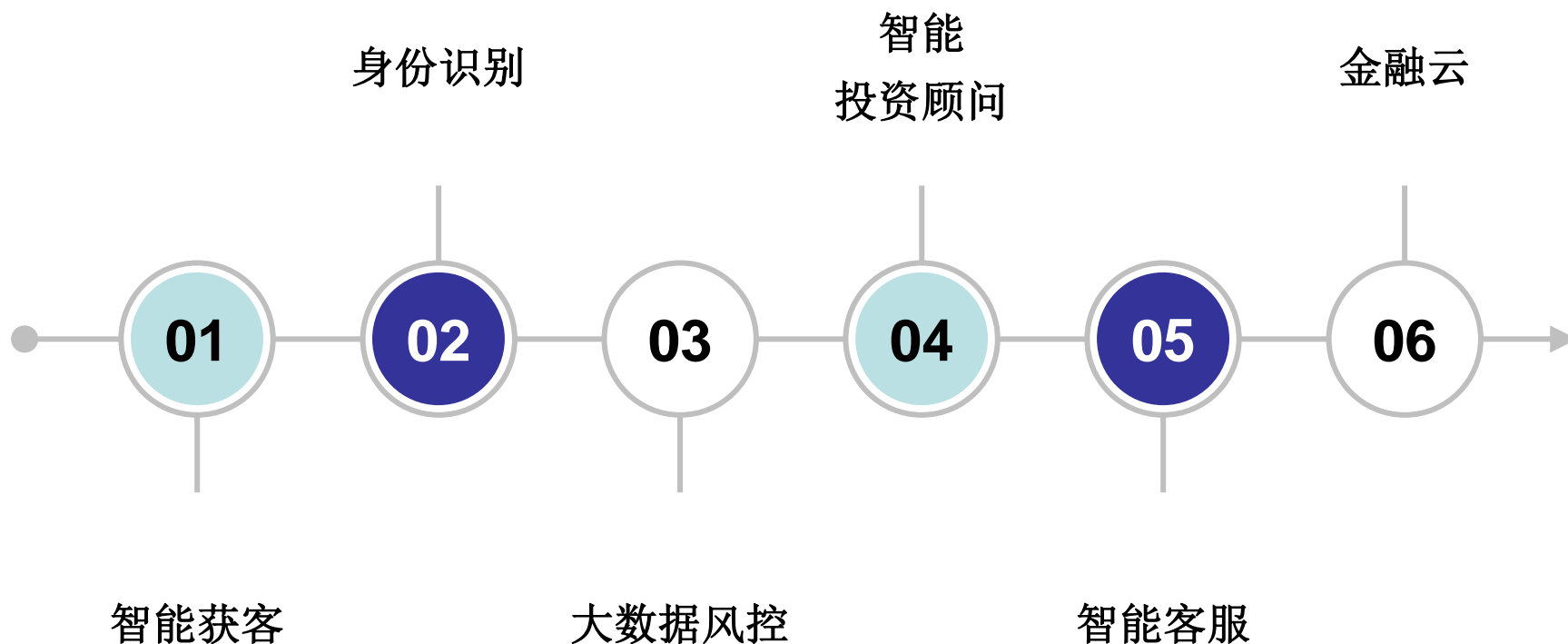
智能家居通过物联网技术将家中的各种设备（如音视频设备、照明系统、窗帘控制、空调控制、安防系统、数字影院系统、影音服务器、影柜系统、网络家电等）连接到一起，提供家电控制、照明控制、电话远程控制、室内外遥控、防盗报警、环境监测、暖通控制、红外转发以及可编程定时控制等多种功能和手段





2.4.3 人工智能的应用

3. 智能金融

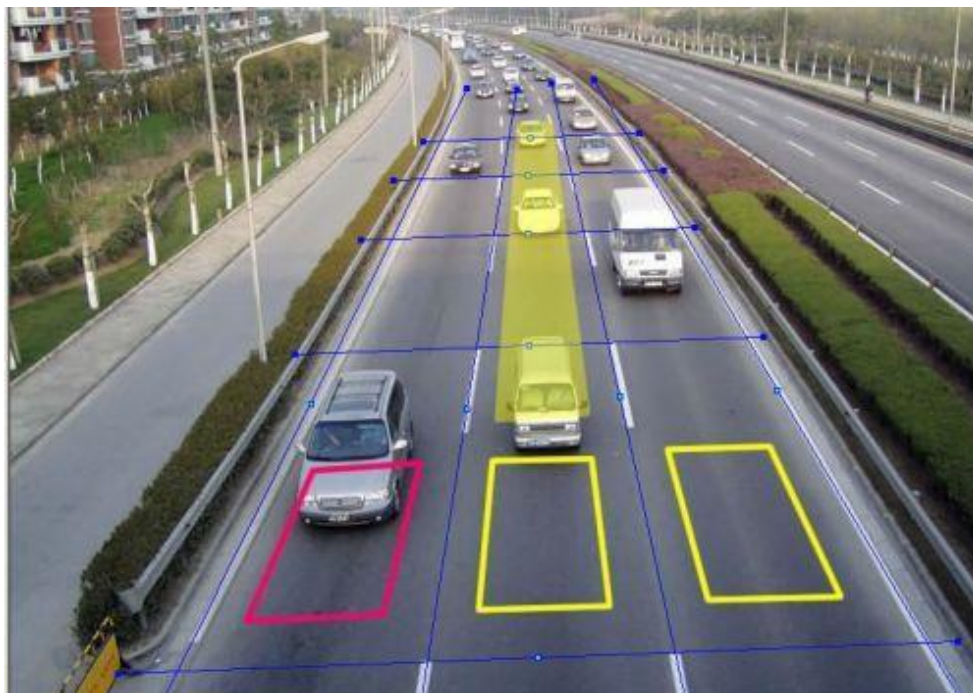




2.4.3 人工智能的应用

4. 智能交通

智能交通是未来交通系统的发展方向，它是将先进的信息技术、数据通讯传输技术、电子传感技术、控制技术及计算机技术等有效地集成运用于整个地面交通管理系统而建立的一种在大范围内、全方位发挥作用的，实时、准确、高效的综合交通运输管理系统。





2.4.3 人工智能的应用

5. 智能安防

- 智能安防是兼顾了整体城市管理系统、环保监测系统、交通管理系统、应急指挥系统等应用的综合体系
- 特别是车联网的兴起，在公共交通安全管理上、车辆事故处理上、车辆偷盗防范上可以更加快捷准确的跟踪定位处理
- 还可以随时随地的通过车辆获取更加精准的灾难事故信息、道路流量信息、车辆位置信息、公共设施安全信息、气象信息等等信息来源



2.4.3 人工智能的应用

6. 智能医疗

借助于物联网/云计算技术、人工智能的专家系统、嵌入式系统的智能化设备，可以构建起完善的物联网医疗体系，使全民平等地享受顶级的医疗服务，解决或减少由于医疗资源缺乏，导致看病难、医患关系紧张、事故频发等现象





2.4.3 人工智能的应用

7. 智能物流

使用智能搜索、推理规划、计算机视觉以及智能机器人等技术，实现货物运输过程的自动化运作和高效率优化管理，提高物流效率



图 京东智能分拣



2.4.3 人工智能的应用

8. 智能零售

人工智能在零售领域的应用已经十分广泛，无人便利店、智慧供应链、客流统计等等都是的热门方向





2.4.4 人工智能产业





2.4.4 人工智能产业

1. 智能基础设施建设





2.4.4 人工智能产业

2. 智能信息及数据

在人工智能数据采集、存储、处理和分析方面的企业主要有两种：

- 一种是数据集提供商
- 另一种是数据采集、存储、处理和分析综合性厂商



2.4.4 人工智能产业

3. 智能技术服务

- 提供人工智能的技术平台和算法模型
- 提供人工智能的整体解决方案
- 提供人工智能在线服务



2.4.4 人工智能产业

4. 智能产品

- 智能机器人
- 智能运载工具
- 智能终端
- 自然语言处理
- 计算机视觉
- 生物特征识别
- VR/AR
- 人机交互



2.4.5 大数据与人工智能的关系

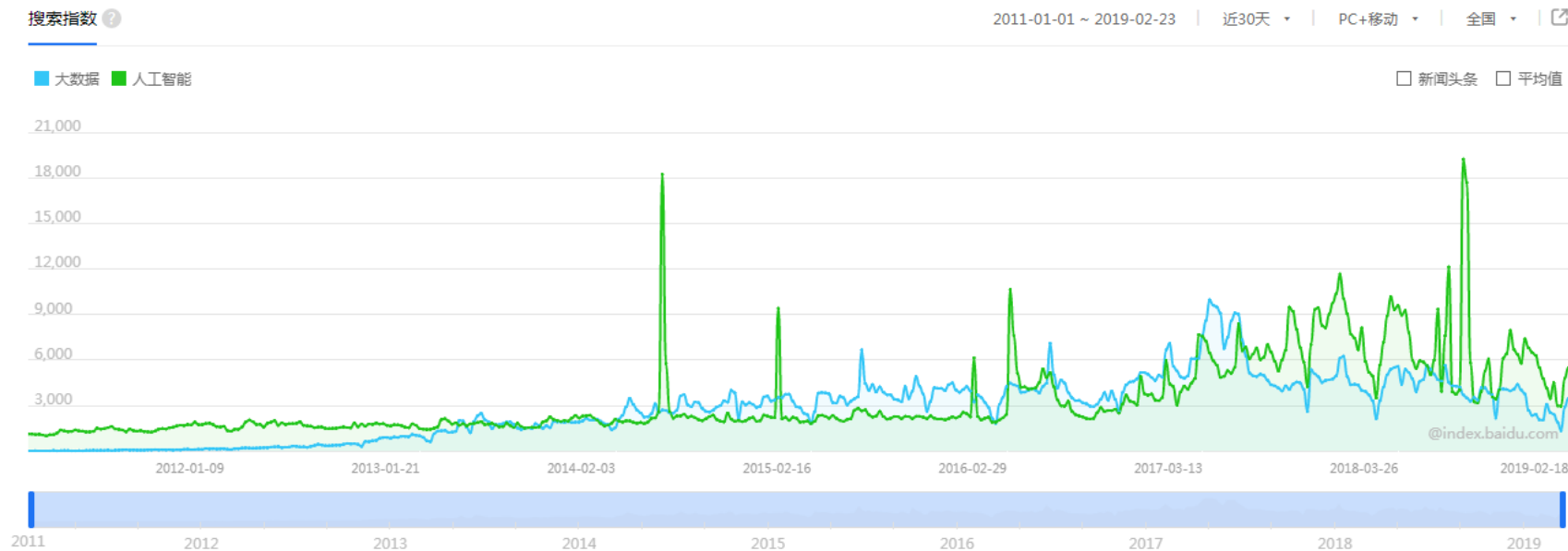


图 人工智能和大数据的百度指数



2.4.5 大数据与人工智能的关系

1.人工智能与大数据的联系

- 人工智能需要数据来建立其智能，特别是机器学习
- 大数据技术为人工智能提供了强大的存储能力和计算能力

2.人工智能与大数据的区别

- 人工智能与大数据也存在着明显的区别，人工智能是一种计算形式，而大数据是一种传统计算，它不会根据结果采取行动，只是寻找结果
- 二者要达成的目标和实现目标的手段不同



2.5 大数据与区块链

2.5.1 从比特币说起

2.5.2 区块链原理

2.5.3 区块链定义

2.5.4 区块链的应用

2.5.5 大数据与区块链的关系



2.5.1 从比特币说起



一切的开始
比特币



2.5.1 从比特币说起

2008年10月31日 中本聪的比特币论文
《比特币:一种点对点的电子现金系统》





2.5.1 从比特币说起

2009年1月3日，在位于芬兰赫尔辛基的服务器上，至今匿名的神秘技术极客中本聪生成了第一个比特币区块，即所谓的比特币“创世区块”





2.5.1 从比特币说起

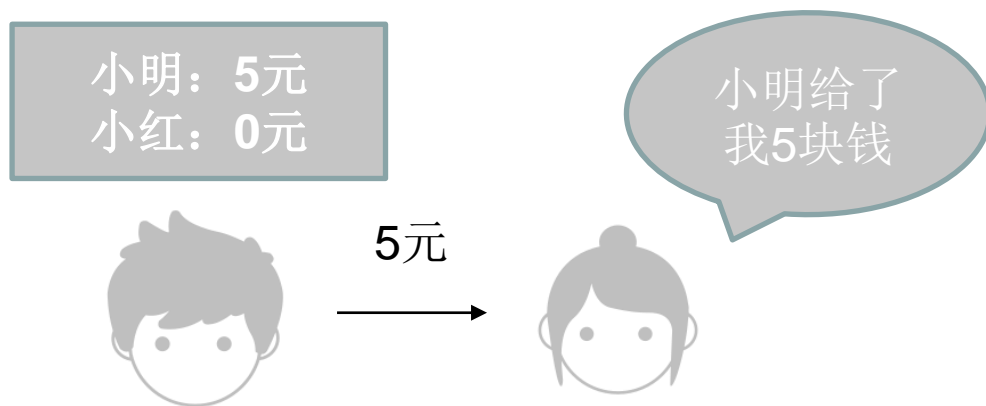
在创世区块的备注中，中本聪写入了当天英国《泰晤士报》的头版头条标题：

The Times 03/Jan/2009 Chancellor on brink of second bailout for banks
“《泰晤士报》，2009年1月3日，财政大臣站在第二次救助银行的边缘”



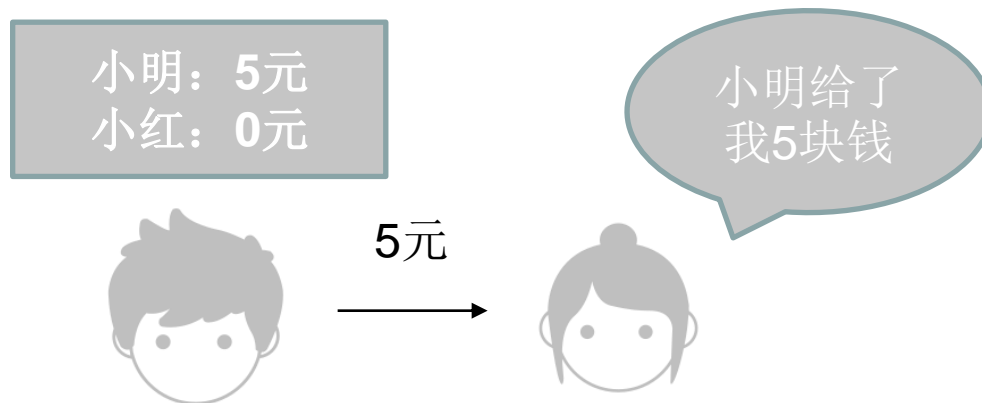
2.5.1 从比特币说起

中本聪为什么要创造比特币





2.5.1 从比特币说起



传统模式:
国家/银行/权威机构负责记账
小明->小红 5块钱

问题:

手续费?

中介系统瘫痪、中介违约、中介欺瞒（民生银行30亿假理财）、
甚至是中介耍赖（塞浦路斯对银行储户一次性征税）？

银行倒闭？被盗？记错帐？

国家货币贬值？超发货币？



2.5.1 从比特币说起



在数字世界中
如何创建一个无须中介或者说去中心化的数字现金
一直是一个难题



2.5.2 区块链原理

1.从记账开始讲起

货币最重要的行为就是交易，交易会产生记录

法币是由我们信任的中心化机构
(政府、银行) 记账

我们的信任往往会被辜负

为了避免以上问题，是否有一种货币可以不用中心化机构来记账，这也是比特币发行的初衷

编号	转账人	收款人	金额
.....
501	王小明	陈云	20
502	张一山	刘大虎	80
503	林彤文	司马鹰松	500
504	李文全	赵明亮	180
505		央行	1000
506	央行	某某	1000
.....



2.5.2 区块链原理

2.比特币要解决的第一个问题：防篡改

不由传统的“可信”的中介机构记账，那么谁来记账？

怎样保证新的记账者不会篡改交易记录呢？

黑客攻击篡改交易记录怎么办？

这就是比特币要解决的第一个问题：防篡改



2.5.2 区块链原理

哈希函数的作用

将任意长度的字符串，转变成固定长度的输出（比如256位），输出的值就被称为哈希值

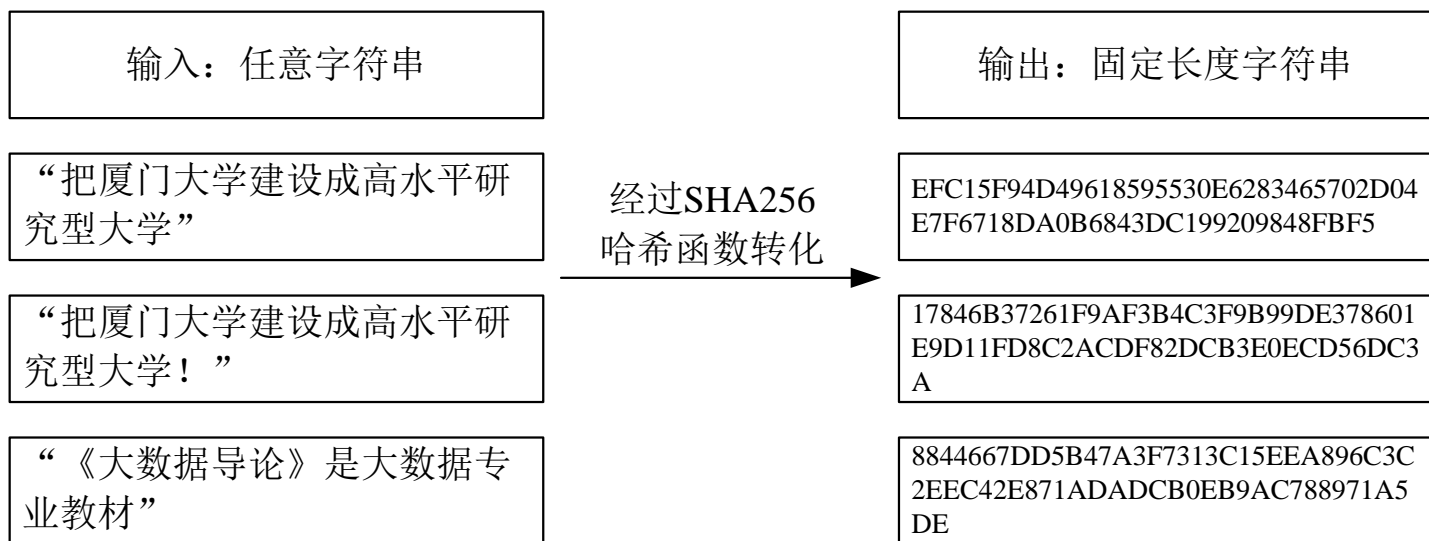


图2-27 使用哈希函数转换的效果

很难找到两个不同的 x 和 y ，使得 $h(x)=h(y)$ ，也就是说，两个不同的输入，会有不同的输出
根据已知的输出，很难找到对应的输入



2.5.2 区块链原理

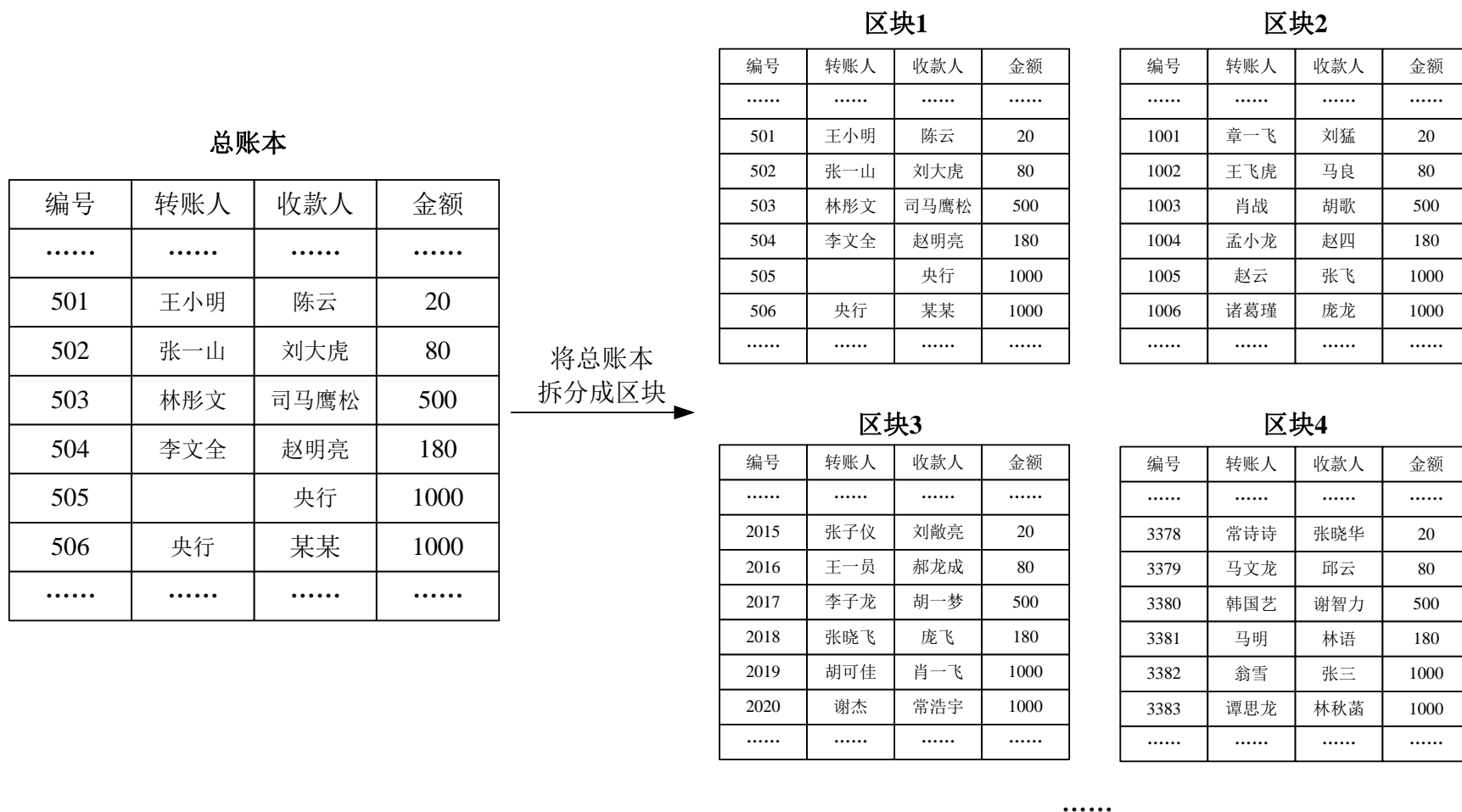


图2-28 把一个大账本拆分成很多个区块



2.5.2 区块链原理

在每个区块上，增加区块头，其中记录了父区块的哈希值
通过每个区块存储父区块的哈希值，把所有区块按照顺序组织起来，形成区块链

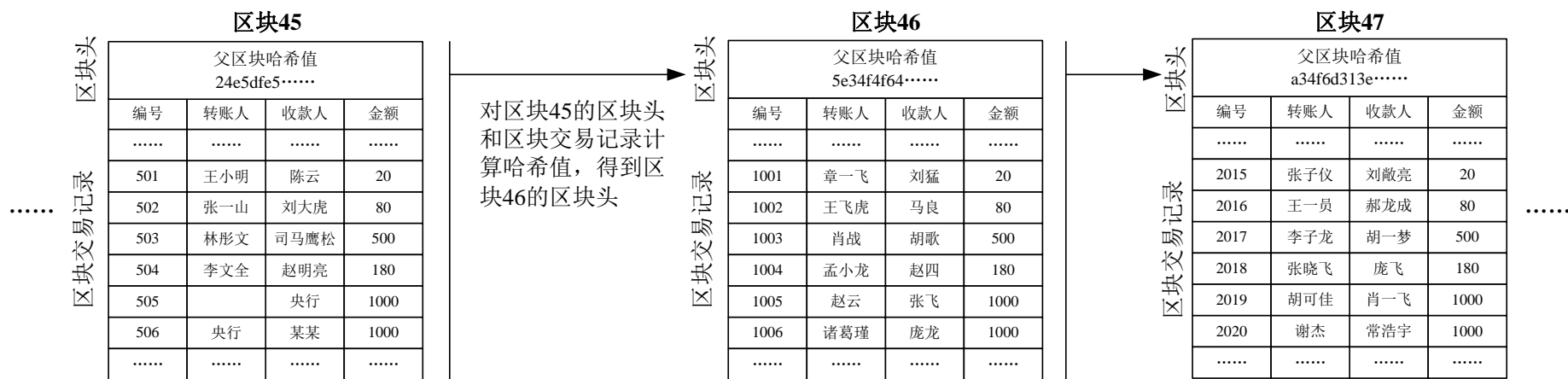


图2-29 区块链示意图



2.5.2 区块链原理

区块链如何防止交易记录被篡改?

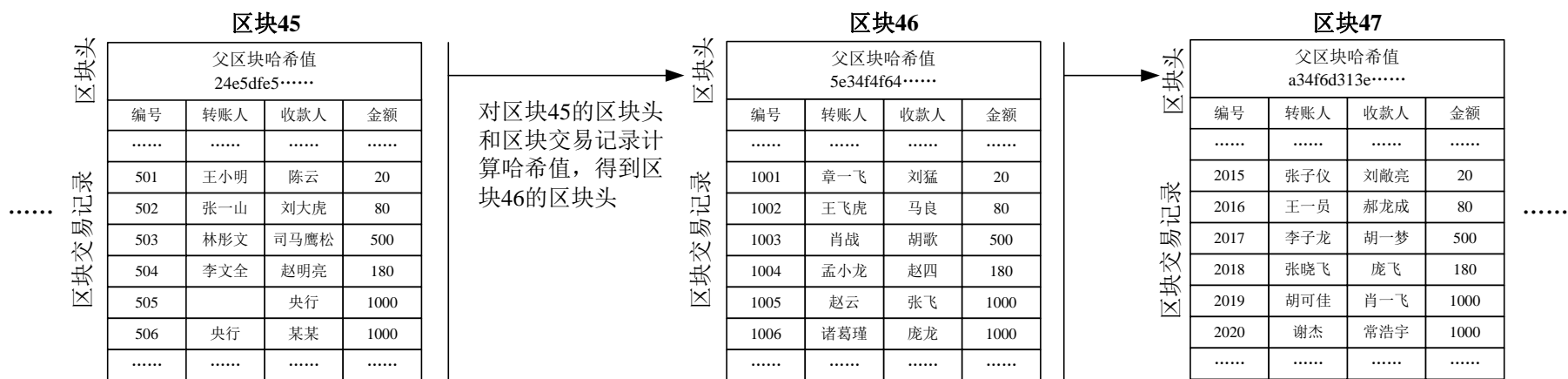


图2-29 区块链示意图

是否可以解决51%攻击问题?

以2017年11月16日的价格计算，在比特币网络进行51%攻击每天的成本包含大约31.4亿美元的硬件成本和560万美元的电力成本



2.5.2 区块链原理

区块链的本质

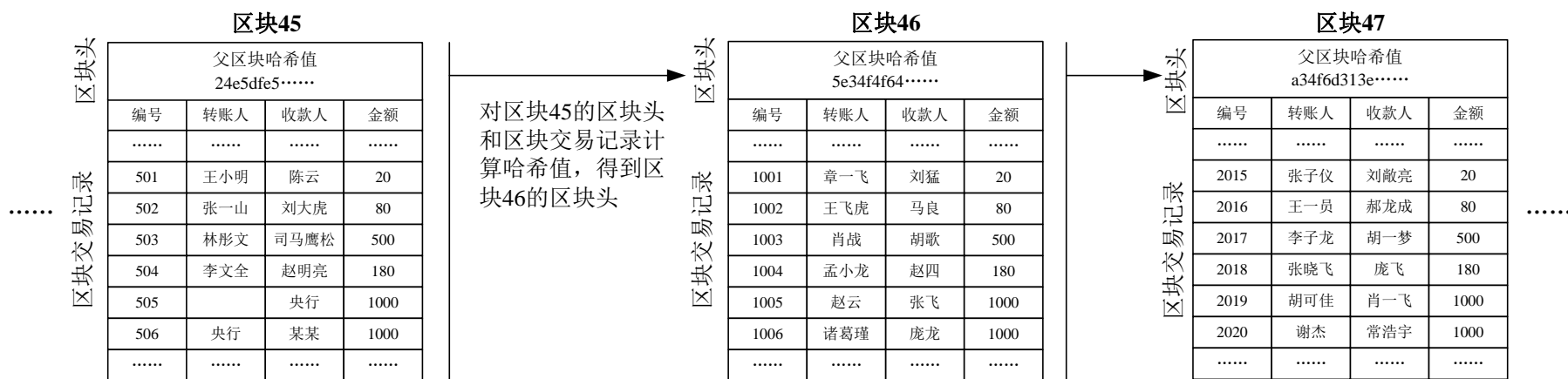


图2-29 区块链示意图

比特币和区块链的本质：就是一个人人可见的大账本，只记录交易

核心技术：通过密码学+数据结构，保证账本记录不被篡改

核心功能：创造信任。法币依靠政府公信力，比特币依靠技术



2.5.2 区块链原理

3、在比特币的世界中如何进行交易

如何交易：地址和私钥

- 进行交易，需要账号和密码，对应公钥和私钥
- 私钥：一串256位二进制数字。获取不需要申请，甚至不需要电脑，可以抛硬币256次来生成
- 地址由私钥转化而成，地址不能反推私钥
- 地址即身份，代表了比特币世界的ID
- 一个地址（私钥）产生以后，只有进入区块链账本，才被大家知道

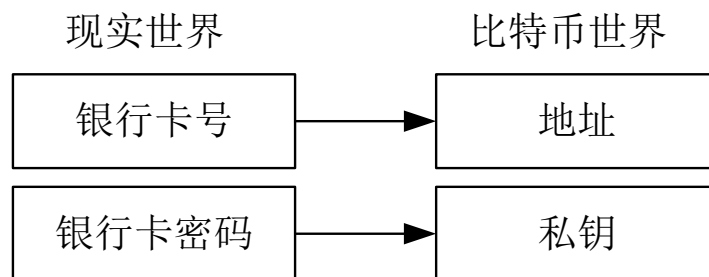


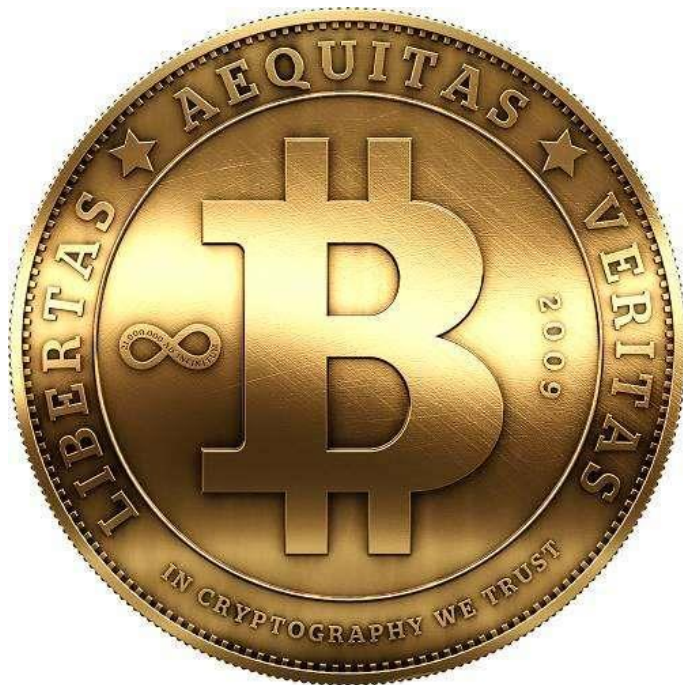
图2-30 银行卡和比特币的对比

两者不同之处：

- 1、银行卡密码可以修改，但是，私钥一旦生成，就无法修改
- 2、银行卡需要申请，而地址和私钥自己就可以生成
- 3、银行卡实名制，地址和私钥是匿名的
- 4、个人申请银行卡有限制，但是地址和私钥可以无限生成



2.5.2 区块链原理



在比特币的世界中，私钥就是一切



2.5.2 区块链原理

数字签名技术

张三已经有了地址、私钥，想要转账给李四10元
如何将这条交易记录添加到区块链中呢？

签名函数 $\text{Sign}(\text{张三的私钥}, \text{转账信息: 张三转10元给李四}) = \text{本次转账签名}$

验证函数 $\text{Verify}(\text{张三的地址}, \text{转账信息: 张三转10元给李四}, \text{本次转账签名}) = \text{True}$

图2-31 签名函数和验证函数

张三通过签名函数 $\text{Sign}()$ ，使用自己的私钥对本次交易进行签名
任何人都可以通过验证函数 $\text{Verify}()$ ，来验证此次签名是否由持有张三私钥的张三本人发出的，是就返回True，反之返回False。
 $\text{Sign}()$ 和 $\text{Verify}()$ 由密码学保证不被破解。



2.5.2 区块链原理

4.比特币要解决的第二个问题：去中心化记账

中心化记账的缺点

然而历史上所有由中心化机构记账的加密数字货币尝试，都失败了

中心化记账的缺点：

- 1、拒绝服务攻击
- 2、厌倦后停止服务
- 3、中心机构易被攻击。比如破坏服务器、网络，监守自盗，法律终止，政府干预

所以比特币需要解决第二个问题：去中心化



2.5.2 区块链原理

去中心化记账：人人都可以记账，每个人都可以保留完整账本。

任何人都可以下载开源程序，参与P2P网络，监听全世界发送的交易，成为记账节点，参与记账





2.5.2 区块链原理

去中心化记账流程

- 1、某人发起一笔交易以后，他向全网广播
- 2、每个记账节点，持续监听、传播全网的交易。收到一笔新交易，验证准确性以后，将其放入交易池，并继续向其他节点传播
- 3、因为网络传播，同一时间，不同记账节点的交易池不一定相同
- 4、每隔10分钟，从所有记账节点当中，按照某种方式抽取1名，将其交易池作为下一个区块，并向全网广播
- 5、其他节点根据最新的区块中的交易，删除自己交易池中已经记录的交易，继续记账，等待下一次被选中



2.5.2 区块链原理

如何分配记账权 (POW机制: 工作量证明机制) 记账节点通过计算一下数学题, 来争夺记账权

找到某随机数, 使得以下不等式成立

SHA256哈希函数 (随机数, 父区块哈希值, 交易池中的交易) < 某一指定值

图2-32 POW机制的数学原理

除了从零开始遍历随机数碰运气以外, 没有其他办法。解题的过程, 又叫挖矿, 记账节点被称为矿工
谁先解对, 谁就获得记账权
某记账节点率先找到解, 就向全网公布, 其他节点验证无误之后, 在新区块之后, 重新开始下一轮计算, 这种方式被称为POW



2.5.2 区块链原理

比特币全貌

区块链（数据结构+哈希函数），保证账本不能被篡改
数字签名技术，保证只有自己才能够动自己的账户
P2P网络和POW共识，保证去中心化的运作方式



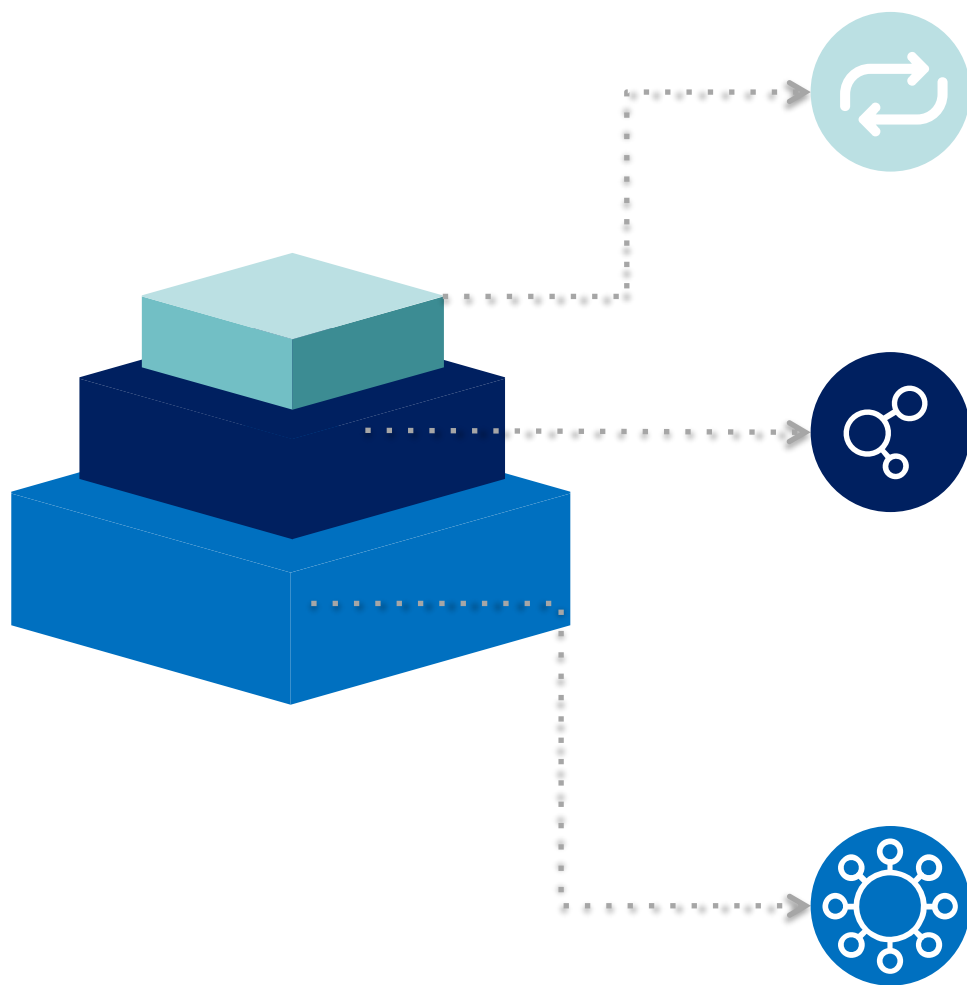


2.5.3 区块链定义

区块链是利用块链式数据结构来验证与存储数据、利用分布式节点共识算法来生成和更新数据、利用密码学的方式保证数据传输和访问安全的一种全新的分布式基础架构与计算范式。



2.5.3 区块链定义



交易 (Transaction)

一次操作，导致账本状态的一次改变，如添加一条记录

区块 (Block)

记录一段时间内发生的交易和状态结果，是对当前账本状态的一次共识；

链 (Chain)

由一个个区块按照发生顺序串联而成，是整个状态变化的日志记录



2.5.4 区块链的应用

从科技层面来看，区块链涉及数学、密码学、互联网和计算机编程等很多科学技术问题。从应用视角来看，简单来说，区块链是一个分布式的共享账本和数据库，具有去中心化、不可篡改、全程留痕、可以追溯、集体维护、公开透明等特点。这些特点保证了区块链的“诚实”与“透明”，为区块链创造信任奠定了坚实的基础。而区块链丰富的应用场景，基本上都基于区块链能够解决信息不对称问题，实现多个主体之间的协作信任与一致行动。

- 金融领域
- 物流领域
- 物联网领域
- 版权保护
- 教育行业
- 数字政务
- 公益和慈善
- 实体资产
- 社交



2.5.5 大数据与区块链的关系

1. 大数据与区块链的区别

(1) 数据量。区块链技术是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式，区块链处理的数据量更小，是细致的处理方式。而大数据管理的是海量数据，要求广度和数量，处理方式上也会更粗糙。

(2) 结构化和非结构化。区块链是结构定义严谨的块，通过指针组成的链，是典型的结构化数据，而大数据需要处理的更多的是非结构化数据。

(3) 独立和整合。区块链系统为保证安全性，信息是相对独立的，而大数据的重点是信息的整合分析。

(4) 直接和间接。区块链是一个分布式账本，本质上就是一个数据库，而大数据指的是对数据的深度分析和挖掘，是一种间接的数据。



2.5.5 大数据与区块链的关系

1. 大数据与区块链的区别

(5) **CAP理论**。**C (Consistency)** 是一致性，它是指任何一个读操作总是能够读到之前完成的写操作的结果，也就是在分布式环境中，多点的数据是一致的。**A (Availability)** 是可用性，它是指快速获取数据，可以在确定的时间内返回操作结果。**P (Tolerance of Network Partition)** 是分区容忍性，它是指当出现网络分区的情况时（即系统中的一部分节点无法和其他节点进行通信），分离的系统也能够正常运行。**CAP理论**告诉我们，一个分布式系统不可能同时满足一致性、可用性和分区容忍性这3个需求，最多只能同时满足其中2个，正所谓“鱼和熊掌不可兼得”。大数据通常选择实现**AP**，而区块链则选择实现**CP**。

(6) **基础网络**。大数据底层的基础设施通常是计算机集群，而区块链则是基于**P2P网络**。

(7) **价值来源**。对于大数据而言，数据是信息，需要从数据中提炼得到价值。而对于区块链而言，数据是资产，是价值的传承。

(8) **计算模式**。在大数据的场景中，是把一件事情分给多个人做，比如，在**MapReduce**计算框架中，一个大型任务会被分解成很多个子任务，分配给很多个节点同时去计算。而在区块链的场景中，是让多个人重复做一件事情，比如，**P2P网络**中的很多个节点同时记录一笔交易。



2.5.5 大数据与区块链的关系

2.大数据与区块链的联系

区块链的可信任性、安全性和不可篡改性，正在让更多数据被释放出来，区块链会对大数据产生深远的影响：

- (1) 区块链使大数据极大降低信用成本
- (2) 区块链是构建大数据时代的信任基石
- (3) 区块链是促进大数据价值流通的管道



2.6 本章小结

云计算、物联网、大数据、人工智能和区块链，代表了人类IT技术的最新发展趋势，五大技术深刻变革着我们的生产和生活。五种技术中，人工智能具有较长的发展历史，在上个世纪五六十年代就已经被提出，并在**2016**年附近迎来了又一次发展高潮。云计算、物联网和大数据在**2010**年附近迎来一次大发展，目前正在各大领域不断深化应用。区块链在**2019**年迅速崛起，引起广泛关注，其应用领域正在不断拓展。本章对云计算、物联网、人工智能和区块链做了简要的介绍，并且梳理了大数据与这四种技术的紧密关系。相信五种技术的融合发展、相互助力，一定会给人类社会的未来发展带来更多的新变化。



附录A：主讲教师林子雨简介



主讲教师：林子雨

单位：厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn

个人网页: <http://dblab.xmu.edu.cn/post/linziyu>

数据库实验室网站: <http://dblab.xmu.edu.cn>



扫一扫访问个人主页

林子雨，男，1978年出生，博士（毕业于北京大学），全国高校知名大数据教师，现为厦门大学计算机科学系副教授，曾任厦门大学信息科学与技术学院院长助理、晋江市发展和改革局副局长。中国计算机学会数据库专业委员会委员，中国计算机学会信息系统专业委员会委员。国内高校首个“数字教师”提出者和建设者，厦门大学数据库实验室负责人，厦门大学云计算与大数据研究中心主要建设者和骨干成员，2013年度、2017年度和2020年度厦门大学教学类奖教金获得者，荣获2019年福建省精品在线开放课程、2018年厦门大学高等教育成果特等奖、2018年福建省高等教育教学成果二等奖、2018年国家精品在线开放课程。主要研究方向为数据库、数据仓库、数据挖掘、大数据、云计算和物联网，并以第一作者身份在《软件学报》《计算机学报》和《计算机研究与发展》等国家重点期刊以及国际学术会议上发表多篇学术论文。作为项目负责人主持的科研项目包括1项国家自然科学基金青年基金项目(No.61303004)、1项福建省自然科学基金青年基金项目(No.2013J05099)和1项中央高校基本科研业务费项目(No.2011121049)，主持的教改课题包括1项2016年福建省教改课题和1项2016年教育部产学协作育人项目，同时，作为课题负责人完成了国家发改委城市信息化重大课题、国家物联网重大应用示范工程区域试点泉州市工作方案、2015泉州市互联网经济调研等课题。中国高校首个“数字教师”提出者和建设者，2009年至今，“数字教师”大平台累计向网络免费发布超过1000万字高价值的研究和教学资料，累计网络访问量超过1000万次。打造了中国高校大数据教学知名品牌，编著出版了中国高校第一本系统介绍大数据知识的专业教材《大数据技术原理与应用》，并成为京东、当当网等网店畅销书籍；建设了国内高校首个大数据课程公共服务平台，为教师教学和学生学习大数据课程提供全方位、一站式服务，年访问量超过200万次，累计访问量超过1000万次。





附录C：林子雨大数据系列教材



林子雨大数据系列教材

用于导论课、专业课、实训课、公共课

了解全部教材信息：<http://dblab.xmu.edu.cn/post/bigdatabook/>



附录D：《大数据导论（通识课版）》教材

开设全校公共选修课的优质教材



本课程旨在实现以下几个培养目标：

- 引导学生步入大数据时代，积极投身大数据的变革浪潮之中
- 了解大数据概念，培养大数据思维，养成数据安全意识
- 认识大数据伦理，努力使自己的行为符合大数据伦理规范要求
- 熟悉大数据应用，探寻大数据与自己专业的应用结合点
- 激发学生基于大数据的创新创业热情

高等教育出版社 ISBN:978-7-04-053577-8 定价：32元

教材官网：<http://dbllab.xmu.edu.cn/post/bigdataintroduction/>



附录E：《大数据技术原理与应用》教材

《大数据技术原理与应用——概念、存储、处理、分析与应用（第2版）》，由厦门大学计算机科学系林子雨博士编著，是国内高校第一本系统介绍大数据知识的专业教材。人民邮电出版社 ISBN:978-7-115-44330-4 定价：49.80元

全书共有15章，系统地论述了大数据的基本概念、大数据处理架构Hadoop、分布式文件系统HDFS、分布式数据库HBase、NoSQL数据库、云数据库、分布式并行编程模型MapReduce、Spark、流计算、图计算、数据可视化以及大数据在互联网、生物医学和物流等各个领域的应用。在Hadoop、HDFS、HBase和MapReduce等重要章节，安排了入门级的实践操作，让读者更好地学习和掌握大数据关键技术。

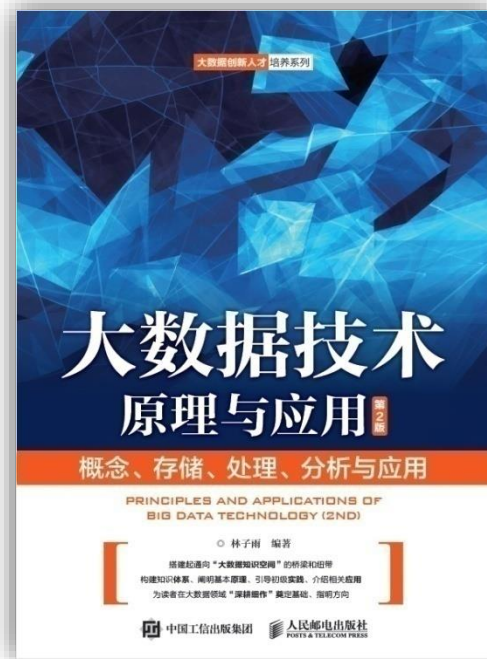
本书可以作为高等院校计算机专业、信息管理等相关专业的大数据课程教材，也可供相关技术人员参考、学习、培训之用。

欢迎访问《大数据技术原理与应用——概念、存储、处理、分析与应用》教材官方网站：

<http://dbllab.xmu.edu.cn/post/bigdata>



扫一扫访问教材官网





附录F：《大数据基础编程、实验和案例教程》

本书是与《大数据技术原理与应用（第2版）》教材配套的唯一指定实验指导书

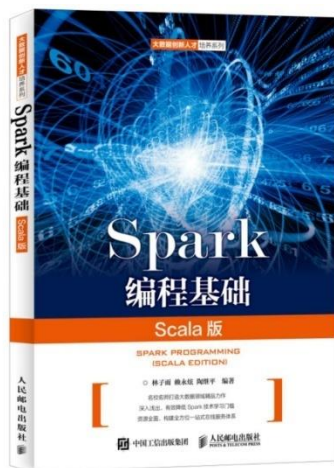


- 步步引导，循序渐进，详尽的安装指南为顺利搭建大数据实验环境铺平道路
- 深入浅出，去粗取精，丰富的代码实例帮助快速掌握大数据基础编程方法
- 精心设计，巧妙融合，五套大数据实验题目促进理论与编程知识的消化和吸收
- 结合理论，联系实际，大数据课程综合实验案例精彩呈现大数据分析全流程

林子雨编著《大数据基础编程、实验和案例教程》
清华大学出版社 ISBN:978-7-302-47209-4 定价：59元



附录G：《Spark编程基础（Scala版）》



《Spark编程基础（Scala版）》

厦门大学 林子雨，赖永炫，陶继平 编著

披荆斩棘，在大数据丛林中开辟学习捷径
填沟削坎，为快速学习Spark技术铺平道路
深入浅出，有效降低Spark技术学习门槛
资源全面，构建全方位一站式在线服务体系

人民邮电出版社出版发行，ISBN:978-7-115-48816-9
教材官网：<http://dblab.xmu.edu.cn/post/spark/>

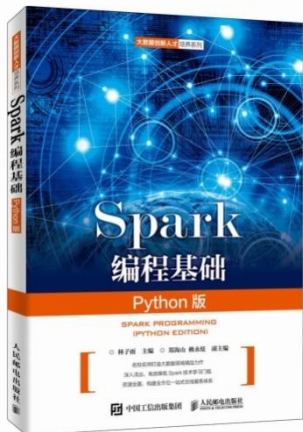


本书以Scala作为开发Spark应用程序的编程语言，系统介绍了Spark编程的基础知识。全书共8章，内容包括大数据技术概述、Scala语言基础、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作，以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源，包括讲义PPT、习题、源代码、软件、数据集、授课视频、上机实验指南等。



附录H：《Spark编程基础（Python版）》

《Spark编程基础（Python版）》



厦门大学 林子雨，郑海山，赖永炫 编著

披荆斩棘，在大数据丛林中开辟学习捷径
填沟削坎，为快速学习Spark技术铺平道路
深入浅出，有效降低Spark技术学习门槛
资源全面，构建全方位一站式在线服务体系

人民邮电出版社出版发行，ISBN:978-7-115-52439-3

教材官网：<http://dbllab.xmu.edu.cn/post/spark-python/>



本书以Python作为开发Spark应用程序的编程语言，系统介绍了Spark编程的基础知识。全书共8章，内容包括大数据技术概述、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Structured Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作，以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源，包括讲义PPT、习题、源代码、软件、数据集、上机实验指南等。



附录I：高校大数据课程公共服务平台



高校大数据课程

公 共 服 务 平 台

<http://dblab.xmu.edu.cn/post/bigdata-teaching-platform/>



扫一扫访问平台主页



扫一扫观看3分钟FLASH动画宣传片



附录J：高校大数据实训课程系列案例教材

为了更好地满足高校开设大数据实训课程的教材需求，厦门大学数据库实验室林子雨老师团队联合企业共同开发了《高校大数据实训课程系列案例》，目前已经完成开发的系列案例包括：

《电影推荐系统》（已经于2019年5月出版）

《电信用户行为分析》（已经于2019年5月出版）

《实时日志流处理分析》

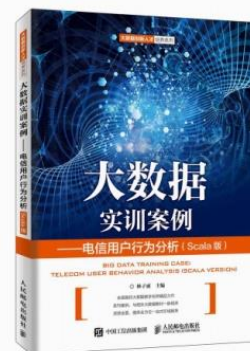
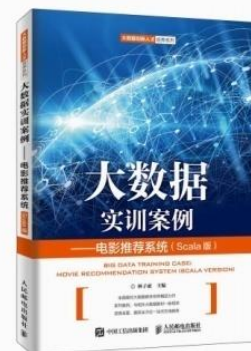
《微博用户情感分析》

《互联网广告预测分析》

《网站日志处理分析》

系列案例教材将于2019年陆续出版发行，教材相关信息，敬请关注网页后续更新！

<http://dblab.xmu.edu.cn/post/shixunkecheng/>



扫一扫访问大数据实训课程系列案例教材主页

The background is a solid blue color. It features several faint, light-blue silhouettes of people. In the top left, a group of people is holding hands in a circle. In the top right, another group of people is standing together. On the right side, there is a large silhouette of a person's head and shoulders. In the bottom left, there is a silhouette of a person's head and shoulders. The text "Thank You!" is centered in the middle of the image in a white, bold, sans-serif font.

Thank You!

Department of Computer Science, Xiamen University, 2020