

目录

一、大数据概述

1、大数据简介

大数据是指数据量巨大、类型繁多、处理速度快的数据集。它的出现是因为互联网和物联网的发展，使得数据的产生和存储成本降低。

2、大数据特征

大数据具有三个特征：数据量大、数据类型多样、数据处理速度快。

3、大数据的技术

等方面，其中包括 Hadoop、Spark、NoSQL 等技术。

4、大数据的应用

大数据的应用非常广泛，包括金融、医疗、电商、智能制造等领域，可以帮助企业做出更明智的决策。

5、大数据处理方法

大数据处理方法包括数据清洗、数据挖掘、数据分析和数据可视化等步骤，可以帮助企业从数据中发现更多的商业价值。

二、大数据发展现状与趋势分析

1、国外现状

、亚马逊和 Facebook 等。这些企业利用大数据来提高产品质量、优化用户体验和提高销售额。

2、国内现状

在国内，大数据的应用也越来越广泛。例如，支付宝利用大数据来防止欺诈和提高用户体验；滴滴出行利用大数据来优化出行路线和提高司机收入。

3、发展趋势分析

未来，大数据的应用将会更加广泛和深入。例如，随着 5G 技术的发展，大数据在智能制造和智慧城市等领域的应用将会越来越广泛。

三、重点应用领域及行业企业分析

1、重点应用领域

能制造等领域。其中，金融领域是大数据应用最为广泛的领域之一，可以帮助银行和保险公司更好地管理风险和提高效率。

2、重点企业

在大数据行业中，有许多知名企业，例如 XXX、XXX、XXX 和 XXX 等。这些企业在大数据技术和应用方面都有着丰富的经验和实践。同时，还有一些创业公司，例如云从科技、DataVisor 和 XXX 等，它们在大数据领域也有着非常突出的表现。

3、国内运营商分析

随着大数据时代的到来，国内运营商也面临着各种挑战。首先，数据量的成倍增长挑战着数据存储能力。其次，数据类型的多样性也挑战着数据挖掘能力。另外，对大数据的处理速度也挑战着数据处理的时效性。最后，数据跨越组织边界传播也挑战着信息安全。因此，国内运营商需要采取相应的对策，以应对这些挑战。

随着大数据时代的到来，相关政策和法规也应运而生。首先，数据生产的相关政策和法规需要制定，以规范数据的生产 and 采集过程。其次，数据共享的相关政策和法规也需要制定，以促进数据共享和利用。最后，隐私保护的相关政策和法规也需要制定，以保护个人隐私和信息安全。

一、大数据概述

1、大数据简介

随着网络和信息技术的不断普及，人类产生的数据量正在呈指数级增长。大量新数据源的出现导致了非结构化、半结构化数据爆发式的增长。这些数据已经远远超越了目前人力所能处理的范畴。因此，大数据的概念应运而生，成为一个新的领域。

2、大数据特征

数据量巨大、多样化、高速度和价值密度低。这些特征使得大数据的处理和分析变得更加复杂和困难，需要采用新的技术和方法来解决。

首先，大数据的体量非常大，起始计量单位至少是 P（1000 个 T）、E（100 万个 T）或 Z（10 亿个 T），远远超过我们所熟知的 G。其次，大数据的类别也非常多样化，来自多种数据源，包括半结构化和非结构化数据。再次，大数据的处理速度非常快，即使在数据体量庞大的情况下也能够实现实时处理。最后，大数据的真实性非常高，这点尤其重要，因为随着新数据源的兴起，传统数据源的局限被打破，信息的真实性和安全性变得更加重要。

大数据技术是指从各种各样类型的巨量数据中，快速获得有价值信息的技术。主要包括数据采集、数据存取、基础架构、数据处理、统计分析、数据挖掘、模型预测和结果呈现等 8 种技术。同时，这些技术形成了批处理、流处理和交互分析三种计算模式。

大数据”的发展将使

零售业净利润增长 以上，制造业的产品开发、组装成本将下降 50%以上。在制造行业，企业可以通过对网上数据分析了解客户需求和市场动向，并对大数据进行分析，从而有效实现对采购和合理库存量的管理，减少因盲目进货而导致销售损失。在商业上，一些国外超市利用对手机的定位和购物推车获得商场内顾客在各处停留时间，利用视频监视图像软件分析顾客购物行为，优化商场布局和货架排列。在政府决策上，分析几十年来的天气数据，将各地降雨、气温、土壤状况和历年农作物产量做成精密图表，就可以预测农产品生产趋势，政府的激励措施、作物存储量和农业服务也可以随之确定。

大数据的处理方法有很多，但普遍适用的大数据处理流程可以概括为四步：采集、导入和预处理、统计和分析，最后是数据挖掘。

采集：大数据采集是指通过多个数据库接收来自客户端的数据，并且用户可以通过这些数据库进行简单的查询和处理工作。传统的关系型数据库 MySQL 和 Oracle 等常用于数据的采集，而 NoSQL 数据库如 Redis 和 MongoDB 也被广泛采用。

采集的主要特点和挑战是并发数高，需要在采集端部署大量数

分片等问题。

导入/预处理：为有效分析海量数据，应将来自前端的数据导入到一个集中的大型分布式数据库或者分布式存储集群，并进行简单的清洗和预处理工作。导入与预处理的特点和挑战是数据量大，每秒钟的导入量经常会达到百兆，甚至千兆级别。

统计分析：利用分布式数据库或者分布式计算集群对存储于其中的海量数据进行普通的分析和分类汇总等，以满足大多数常见的分析需求。实时性需求可以使用 EMC 的 GreenPlum、XXX 的 Exadata 以及基于 MySQL 的列式存储 Infobright 等，批处理或基于半结构化数据的需求可以使用 Hadoop。统计分析的主要特点和挑战是分析涉及的数据量大，对系统资源，特别是 I/O 会有极大的占用。

挖掘：数据挖掘是在现有数据上进行基于各种算法的计算，从而起到预测的效果，实现高级别数据分析的需求。常用算法有 Kmeans、SVM 和 NaiveBayes，主要工具有 Hadoop 的

量和计算量都很大，常用数据挖掘算法都以单线程为主。

国外现状：在国外，大数据的应用已经非常广泛，尤其在金融、医疗、电商等领域。大数据技术不断发展，越来越多的企业开始采用云计算和深度研究等技术来提高数据处理和分析的效率。同时，人工智能、物联网和区块链等技术也将与大数据技术相结合，为企业提供更多的数据处理和分析方案。

趋势分析：未来，大数据技术将继续发展，越来越多的企业将采用大数据技术来处理和分析数据。同时，随着人工智能和物联网技术的发展，大数据技术将与这些技术相结合，为企业提供更多的数据处理和分析方案。在未来的发展中，大数据技术将更加注重数据隐私和安全，同时也需要更加智能化和自动化的数据处理和分析方案。

从 2.6 亿元增长到 11.2 亿元，预计未来几年将迎来高速增长期，有望突破 40 亿元。虽然国内起步较晚，但发展迅速。政府也在积极介入推动大数据产业的发展。2015 年，国务院印发《促进大数据发展行动纲要》，提出了大数据发展的目标和重点领域。同时，资本市场也对大数据产业表现出浓厚的兴

年，XXX、XXX、XXX 等互联网巨头纷纷布局大数据，成为国内大数据产业的主要推动力量。人才需求也在不断增加，对于深度数据分析和管理的的人才需求更是巨大。

总体来看，大数据产业在全球范围内都呈现出快速增长的势头。政府、资本市场和企业都对大数据产业表现出浓厚的兴趣和投资热情。在国内，虽然起步较晚，但发展速度也非常迅速，未来几年有望迎来高速增长期。同时，人才需求也在不断增加。大数据产业已经成为 IT 领域的重要新兴产业，具有广阔的发展前景和巨大的市场潜力。

自 2011 年以来，XXX 和 XXX 先后成立了 XXX，专注于研究大数据中的科学和工程问题。同时，XXX 的《XXX 发展十二五”专项规划》和 XXX 的《物联网“十二五”发展规划》也将大数据技术作为重点支持的领域之一。XXX 的规划中，信息处理技术被提出作为四项关键技术创新工程之一，其中包括海量数据存储、数据挖掘、图像视频智能分析等，这些都是大数据的重要组成部分。另外三项关键技术创新工程，包括信息感知技术、信息传输技术和信息安全技术，也与大数据密切相关。

运营商积极推进大数据应用工作，并取得了较好的进展。电子商务企业 XXX 提出要做中国数据分析第一平台，通过掌握的企业交易数据，借助大数据技术自动分析判定是否给予企业贷款，全程不会出现人工干预。据透露，截至目前 XXX 已经放贷 300 多亿元，坏账率约 0.3% 左右，低于商业银行。

虽然我国处理大数据的企业并不是很多，但是一些走在大数据行业前列的企业，如 XXX，在大数据、分布式计算、数据分析等领域具备核心竞争力，自主创新并拥有多项发明专利。推出的 Z 系列产品在大数据的应用分析中在国际上也是领先的。此外，还有其他一些大数据相关的上市公司，如 XXX、XXX、XXX、XXX、XXX、XXX、XXX、久其软件、XXX 软件、XXX、XXX、XXX、汉得信息、卫士通和 XXX 星辰等。

大数据的热潮也引发了一场思想启蒙运动，使得“大数据是资产，不是包袱”、“要拿数据说话”等观念逐步深入人心，改变了以往不重视数据积累、不相信数据分析等认识。有了这种思维模式的改变，大数据的应用就有了希望。

2011 年全球被创建和被复制的数据总量达 1.8ZB, 预计到 2020 年, 全球电子设备存储的数据将增至 35.2ZB。人类社会正在加速跨进大数据时代, 这是继蒸汽时代、电气时代和网络时代后的又一个重要时代。大数据的发展趋势和阶段正在不断演化, 开源软件和产业垂直整合是其中的两个重要方向。

在大数据时代, 企业对于解决业务问题的需求越来越迫切, 因此大型企业越来越倾向于采用定制化解决方案来满足自身的需求。这些定制化解决方案通常由大型软件厂商或者专业的解决方案提供商提供, 他们会根据企业的需求, 量身定制出一套适合该企业的解决方案。这些解决方案往往包括了软件、硬件、服务等多个方面, 以满足企业的特定需求。这种定制化解决方案的出现, 也进一步推动了大数据产业的发展, 促进了企业的数字化转型。

5)、人工智能技术的快速发展

人工智能技术是大数据时代的另一个重要方向。随着深度研究等技术的快速发展, 人工智能在图像识别、语音识别、自然语言处理等领域取得了很大的进展。在大数据时代, 人工智

效率和竞争力。目前，国内外的很多大型企业都已经开始大力投入人工智能技术的研发和应用，这也进一步推动了人工智能技术的发展。

6)、大数据安全问题成为新的挑战

随着大数据的快速发展，大数据安全问题也成为了一个新的挑战。大数据中包含了很多敏感信息，例如个人隐私、商业机密等，这些信息一旦泄露将会对企业带来很大的损失。因此，大数据安全问题也成为了企业必须要面对的一个问题。目前，国内外的很多企业都已经开始加强大数据安全的管理和保护，投入更多的资源来解决这个问题。同时，也有很多安全厂商开始提供大数据安全解决方案，帮助企业更好地保护自己的数据安全。

政府、医疗、教育、城市管理等。大数据技术可用于政府决策、医疗诊断、教育评估、城市规划等方面，提高公共服务水平和效率。

于风险控制、客户画像、营销策略等方面，提高金融机构的竞争力。

3)、制造业领域：汽车、航空、机械等。大数据技术可用于生产调度、质量控制、售后服务等方面，提高制造业的效率和品质。

4)、电商领域：电商平台、物流配送等。大数据技术可用于用户画像、精准营销、物流优化等方面，提高电商企业的竞争力和用户体验。

2、行业企业分析

随着大数据技术的发展，越来越多的企业开始关注和应用大数据。国内外已经出现了众多以大数据技术为核心的企业，如 Cloudera、XXX、Google 等。国内的 BAT（XXX、XXX、XXX）也在大数据领域有所布局。此外，各行各业的企业也在积极探索大数据技术的应用，如 XXX、金融机构、汽车制造商等。大数据已经成为企业提高效率、优化业务、创造价值的重要手段。

3、大数据产业链全景图

数据库、数据仓库、数据转换工具、数据安全、组织与管理层、数据挖掘、数据统计、
、AI、内容/知识管理、数据可视化、数据租售、分析与预测服务、决策支持服务、数据分享平台、数据分析平台、IT 基础设施、存储设备、运算设备、一体机、基础软件、虚拟化、中间件、并行运算、数据安全、IT 服务、DAS、NAS、SAN、磁带库、PC 服务器、小型机、大型机、HPC、专家系统、大数据一体机、集装箱数据中心、规划咨询、系统集成、云存储等多个环节。这些环节相互依存、相互支撑，构成了大数据产业链。

据 XXX 的统计分析，我国公共领域对大数据的应用主要集中在电力、智能交通、电子政务和司法系统四个方面。

在电力行业，大数据主要应用于智能电网上，通过获取人们的用电行为信息，智能电网能够优化电的生产、分配和消耗，从而有利于电网安全监测和控制，客户用电行为分析以及客户细分。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/166053054001010123>