大数据
基础与管理

Big Data
Foundation and Management

段竹 田宏 主编

清华大学出版社

大数据
基础与管理

Big Data
Foundation and Management

段竹 田宏 主编
吴旭东 吴镝 冯瑞芳 朱毅 于书皓 编著

清华大学出版社
北京

内 容 简 介

本书从理论结合实践的角度，讲解大数据的概念和技术。全书共分为7章，主要内容包括什么是大数据、大数据的特征、大数据的作用与应用、大数据的技术与分析；通过实例讲解Data Studio的使用方法，DB2、UDB和JDBC的相关知识，集成数据管理的知识，IBM InfoSphere软件；详细讨论大数据环境下的安全与治理；通过实例讲解了Hadoop技术。

图书在版编目(CIP)数据

大数据基础与管理/段竹，田宏主编. --北京：清华大学出版社，2016
ISBN 978-7-302-42523-6

Ⅰ. ①大…　Ⅱ. ①段… ②田…　Ⅲ. ①数据处理 ②数据管理　Ⅳ. ①TP274

中国版本图书馆CIP数据核字(2016)第000889号

责任编辑：刘向威
封面设计：文　静
责任校对：焦丽丽
责任印制：杨　艳

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
课 件 下 载：http://www.tup.com.cn，010-62795954

印 装 者：北京鑫海金澳胶印有限公司
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张**：12.25　　**字　　数**：300千字
版　　次：2016年3月第1版　　**印　　次**：2016年3月第1次印刷
印　　数：1～2000
定　　价：29.00元

产品编号：057643-01

FOREWORD 前言

大数据作为一种重要的战略资产，已经不同程度地渗透到各个行业领域和部门，其深度应用不仅有助于企业经营活动，还有利于推动国民经济发展。大数据的快速发展使它成为IT领域的又一大新兴产业。

大数据目前在很多的行业都有应用，包括大科学(Megascience)、RFID、感测设备网络、天文学、大气学、基因组学、生物学、大社会数据分析、互联网文件处理、制作互联网搜索引擎索引、通信记录明细、军事侦察、社交网络、通勤时间预测、医疗记录、照片图像和视频封存、大规模电子商务等。

对大数据的处理分析正在成为新一代信息技术融合应用的节点。移动互联网、物联网、社交网络、数字家庭、电子商务等是新一代信息技术的应用形态，这些应用不断产生大数据。云计算为这些海量、多样化的大数据提供存储和运算平台。通过对不同来源数据的管理、处理、分析与优化，将结果反馈到上述应用中，创造出巨大的经济和社会价值。

因此，了解大数据的概念，掌握与大数据相关的技术，对于计算机专业的学生来说是十分必要的。

本书从理论结合实践的角度，讲解大数据的概念和技术。读者从本书可以了解到什么是大数据、大数据的特征、大数据的作用与应用、大数据的技术与分析、集成数据管理、大数据环境下的安全与治理、大数据相关技术的使用等知识。

本书共分为7章。

第1章介绍什么是大数据、大数据的特征、大数据的作用与具体应用、大数据的技术与大数据的分析。

第2章通过实例详细讲解Data Studio的使用方法。

第3章介绍DB2与JDBC支持，讲解DB2、UDB和JDBC的通用驱动程序。

第4章详细讲解集成数据管理的知识，包括集成数据管理的基本知识、数据建模和设计、数据模型管理、主数据管理、元数据管理、数据的交付。

第5章详细讲解IBM InfoSphere软件。

第6章详细讲解大数据环境下的安全与治理，包括大数据环境下的信息安全、大数据面临的安全问题、大数据安全的应对策略、大数据的治理、大数据加密技术。

第7章通过实例详细讲解Hadoop技术。

本书不足之处恳请读者指正。

编　者

2015年9月

CONTENTS 目录

第1章 大数据概论

1.1 什么是大数据

大数据(Big data),或称巨量数据、海量数据,指的是所涉及的数据量的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。

1.2 大数据的特征

目前较为普遍的大数据定义为"大数据主要指无法使用传统流程或工具处理和分析的数据"。IBM 所称的大数据通常用 3 个特征描述:数量(Volume)、种类(Variety)和速度(Velocity)。2012 年,英特尔大数据论坛上,IDC 定义了大数据的四大特征:海量的数据规模;快速的数据流转和动态的数据体系;多样的数据类型;巨大的数据价值。综合上述定义,可以用大量(Volume)、多样性(Variety)、速度快(Velocity)以及价值高和密度低(High Value and Low Density)四大特征来描述大数据。

(1) 大量化(Volume)。数据量级已从 GB 至 TB、PB 乃至 ZB 上升,可称海量、巨量甚至超量,且仍在持续爆炸式增长。据 WinterCorp 调查显示,最大数据仓库中的数据量,每两年增加 3 倍左右,其增长速度远超摩尔定律增长速度。谷歌公司每天要处理超过 24PB 的数据,Facebook 每天更新的照片量超过 1000 万张,Twitter 每天都会发布超过 4 亿条微博。截止到 2012 年 12 月底,中国网页数量为 1227 亿个左右,比 2011 年同期增长 41.7%左右。大数据环境下,网络信息的规模急剧增长,PB 级甚至 ZB 级的数据需要大规模并行计算网络的支持,巨大的存储、链接、传输和加密归并等开销使常规加密计算不堪重负。

(2) 多样化(Variety)。数字信息由原来简单的数值、字符和文本向网页、图片、视频、图像和位置信息等半结构化和非结构化的数据类型发展,并且信息大多分布在不同的地理位置、不同的存储设备以及不同的数据管理平台。此外,互联网环境中大量的信息缺乏有效的组织,信息的无序化大幅度降低了查找和利用信息的效率,阻碍了有价值信息的加密和管理效率。具体体现在如下 3 个方面:

- 数据来源多。随着互联网和物联网技术的飞速发展，可以通过微博、社交网站、电子商务网站、车载网以及遍布全球的各式各样的传感器等多种数据来源获取数据。
- 数据类型繁多。传统数据大多以表格的形式保存，而大数据中70%～85%的数据是图片、音频、视频、网络日志、链接信息等半结构化和非结构化的数据。
- 数据之间关联性强，交互频繁。如大型的电子商务网站和社交网络中，一些用户的点击行为在一定程度上反映了该用户潜在的兴趣爱好和需求，链接之间的关联性较强。

(3) 快速化(Velocity)。大数据的时效性要求对数据的处理能够做到实时、快速，要达到这一目标，要求使用的硬件平台亦能够同步更新换代，并将分布式计算、并行计算、软件工程、人工智能等技术应用到其中。

(4) 价值高和密度低(High Value and Low Density)。互联网充斥着大量重复和虚假信息，通常有价值的信息较为分散，密度很低。

正是大数据的价值具备稀疏性、多样性和不确定性的特点，较多数据采集和存储系统又要求能够快速访问大数据的历史版本数据，备份数据的保存期限更长，备份的窗口不断缩短，很多数据需要在线备份和故障实时恢复等，大数据的安全维护对存储资源、计算资源、网络资源等都提出了极高的性能需求，其安全存储与数据保护面临着前所未有的压力和挑战。

1.3 大数据的作用与具体应用

1.3.1 大数据的作用

(1) 对大数据的处理分析正在成为新一代信息技术融合应用的节点。移动互联网、物联网、社交网络、数字家庭、电子商务等是新一代信息技术的应用形态，这些应用不断产生大数据。云计算为这些海量、多样化的大数据提供存储和运算平台。通过对不同来源数据的管理、处理、分析与优化，将结果反馈到上述应用中，将创造出巨大的经济和社会价值。大数据具有催生社会变革的能量。但释放这种能量，需要严谨的数据治理、富有洞见的数据分析和激发管理创新的环境(Ramayya Krishnan，卡内基·梅隆大学海因兹学院院长)。

(2) 大数据是信息产业持续高速增长的新引擎。面向大数据市场的新技术、新产品、新服务、新业态会不断涌现。在硬件与集成设备领域，大数据将对芯片、存储产业产生重要影响，还将催生一体化数据存储处理服务器、内存计算等市场。在软件与服务领域，大数据将引发数据快速处理分析、数据挖掘技术和软件产品的发展。

(3) 大数据利用将成为提高核心竞争力的关键因素。各行各业的决策正在从“业务驱动”转变为“数据驱动”。

对大数据的分析可以使零售商实时掌握市场动态并迅速做出应对；可以为商家制定更加精准有效的营销策略提供决策支持；可以帮助企业为消费者提供更加及时和个性化的服务；在医疗领域，可提高诊断准确性和药物有效性；在公共事业领域，大数据也开始发挥促进经济发展、维护社会稳定等方面的重要作用。

(4) 大数据时代科学研究的方法手段将发生重大改变。例如，抽样调查是社会科学的基本研究方法。在大数据时代，可通过实时监测、跟踪研究对象在互联网上产生的海量行为

数据，进行挖掘分析，揭示出规律性的东西，提出研究结论和对策。

1.3.2　大数据的应用

大数据目前在很多的行业都有应用，包括了大科学(Mega Science)、RFID、感测设备网络、天文学、大气学、基因组学、生物学、大社会数据分析、互联网文件处理、制作互联网搜索引擎索引、通信记录明细、军事侦察、社交网络、通勤时间预测、医疗记录、照片图像和视频封存、大规模的电子商务等。下面将给出几个具体的大数据应用案例供大家参考。

(1) 大科学

大科学领域的一个典型代表是大型强子对撞机(Large Hadron Collider)，大数据在大科学中有着广泛的应用。图 1.1 所示为大型强子对撞机的一部分。大型强子对撞机中有 1.5 亿个感测器，每秒发送 4000 万次的数据。实验中每秒产生将近 6 亿次的对撞，在过滤去除 99.999%的撞击数据后，得到约 100 次的有用撞击数据。将撞击结果数据过滤处理后仅记录了 0.001%的有用数据，四个对撞机的全部数据量复制前每年产生 25 拍字节(PB)，复制后为 200 拍字节。

如果将所有实验中的数据在不过滤的情况下全部记录，数据量将会变得过度庞大且很难处理。每年数据量在复制前估计将会达到 1.5 亿拍字节，等于每天有近 500 艾字节(EB)的数据量。这个数字代表每天实验将产生相当于 500 垓(5×10^{20})字节的数据，是全世界所有数据源总和的 200 倍左右，在如此庞大的数据中去寻找希格斯玻色子(Higgs boson)存在的证据，就需要借助大数据的应用了。

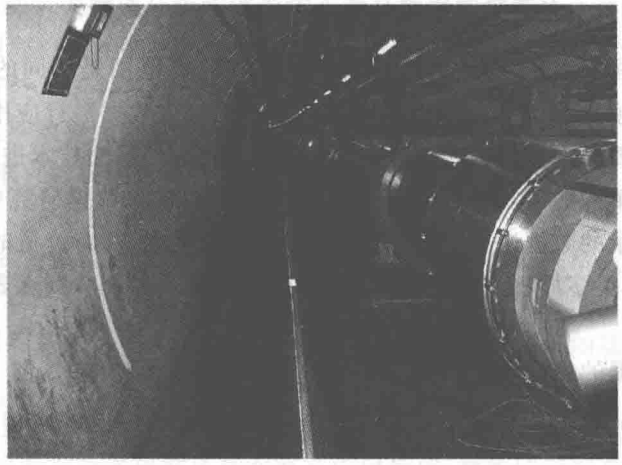

图 1.1　大型强子对撞机的一部分

(2) 社会学

大数据产生的背景离不开微博、Facebook 等社交网络的兴起，人们每天通过这种自媒体传播信息或者沟通交流，由此产生的信息被网络记录下来，社会学家可以在这些数据的基础上分析人类的行为模式、交往方式等。美国的涂尔干计划就是依据个人在社交网络上的数据分析其自杀倾向，该计划从美军退役士兵中拣选受试者，透过 Facebook 的行动应用收集资料，并将用户的活动数据传送到一个医疗资料库。收集完成的数据会接受人工智能系统分析，接着利用预测程序来即时监视受测者是否出现一般认为具有伤害性的行为。

(3) 营销与销售

企业与客户之间的关系早在第一次销售之前就已建立。大数据营销技术可以帮助企业在客户仍然研究自己的选择时与其客户接触,并证明能够提供客户需要的个性化产品和服务。如图 1.2 所示,个人感觉和真实数据的结论往往不同。

如果要赢得客户,需要单独了解客户。需要充分利用企业内外的所有数据,对客户的需求和行为进行智能预测,提高客户保留率和忠诚度。通过更好地了解客户,企业可以提供适当的服务水平,从而提高客户保留率,将客户转变为支持者。由此可见大数据可以帮助企业在营销与销售中取得优秀的成果。

图 1.2 个人感觉和真实数据的结论往往是不同的

1.4 大数据的技术与大数据的分析

1.4.1 概述

在不同的行业领域当中,对大数据的具体运用和研究范围是不同的,普通的行业用户,例如金融、商业、遥感等领域的用户仅仅需要在应用层对大数据进行应用。而计算技术的研究人员和开发人员要对大数据技术在算法层、系统层和平台层的技术进行研究和开发。

大数据不单是单纯的事实存在的很大数量的数据。在大数据研究中,一个很重要的问题是如何对这些海量的数据进行分析,并从中得到一些有意义、有价值的数据。大数据的数据体量巨大、数据种类多样、价值密度低和要求处理速度快的四个特点表现了大数据分析不断增加的复杂性。

本节将对大数据的技术和大数据的分析进行简要的介绍。

1.4.2 大数据的技术

不同的用户对大数据技术的运用是不同的,对大数据的研究领域也不尽相同。本节将对不同层次的大数据技术进行介绍。

大数据的最顶层为应用层,这一层的研究层次包含大数据行业应用/服务层与应用开发层。在这个层次上,大数据需要的技术有行业应用系统和服务的使用与开发,还有领域应用、服务需求和计算模型的制定技术等。

第二层为算法层,包括应用算法层与基础算法层。

在应用算法层,大数据主要的研究内容为社会网络,商业智能,推荐系统,自然语言处理与 Web 数据的挖掘与搜索等。在每个研究内容中又会细分为很多项具体技术,由于篇幅限

制，在这里不多做介绍。

在基础算法层，对于大数据的主要研究内容为大数据并行化机器学习和挖掘方法。具体的算法包括分类算法（Classification）、聚类（Clustering）和参数估计（Parameters Estimation）等。

第三层为系统层，该层包括了并行编程模型与计算框架层和大数据存储管理层。

在并行编程模型与计算框架层，研究领域包括了 MapReduce 在不同构架上实现的技术，定制式并行计算框架与 CUDA、MPI 等技术。

在大数据存储管理层，大数据的研究领域包括大数据预处理技术，记录型大数据索引和查询技术，SQL/NoSQL 查询语言接口技术，分布式数据库与分布式文件系统技术等。

最底层为平台层，该层只有一层内容即并行构架和资源平台层。该层的技术主要有共享内存构架技术、分布内存构架技术、混合式构架技术与大数据应用/服务云计算支撑平台技术等。

1.4.3 大数据的分析

在大数据的研究中，如何从海量的数据中分析并获取到有价值的数据是一个非常重要的问题。大数据分析主要有五个方面，本节将介绍一些大数据分析的方法。

（1）可视化分析。大数据分析的使用者有大数据分析专家，同时还有普通用户，但是他们二者对于大数据分析最基本的要求就是可视化分析。因为可视化分析能够直观地呈现大数据的特点，同时能够非常容易被用户所接受，就如同看图说话一样简单明了。

（2）数据挖掘算法。大数据分析的理论核心就是数据挖掘算法。各种数据挖掘的算法基于不同的数据类型和格式才能更加科学地呈现出数据本身具备的特点，也正是因为这些被全世界统计学家所公认的各种统计方法才能深入数据内部，挖掘出公认的价值。另外一个方面也是因为有这些数据挖掘的算法才能更快速地处理大数据，如果一个算法要花上好几年才能得出结论，那大数据的价值也就无从说起了。

（3）预测性分析。大数据分析的应用领域之一就是预测性分析，从大数据中挖掘出特点，通过科学建模之后便可以通过模型带入新的数据，从而预测未来的数据。

（4）语义引擎。非结构化数据的多元化给数据分析带来新的挑战，这就需要一套工具能够系统地分析和提炼数据。语义引擎需要有足够的人工智能系统的功能以便从数据中主动地提取信息。

（5）数据质量和数据管理。大数据分析离不开数据质量和数据管理，高质量的数据和有效的数据管理，无论是在学术研究还是在商业应用领域，都能够保证分析结果的真实和有价值。

总之，以上五个方面是大数据分析的基础，如果想对大数据进行更加深入的分析，那么需要一些更加独特和专业的方法对数据进行分析。

1.4.4 大数据现状

据一则大数据发展分析报告称，大数据的快速发展，使它成为 IT 领域的又一大新兴产业。据中央财经大学中国经济管理研究院博士张永力估算，国外大数据行业约有 1000 亿美元的市场，而且每年都以大约 10%的速度在增长，增速是软件行业的两倍。

1. 政府积极介入推动

2009 年，联合国启动“全球脉动计划”，借大数据推动落后地区发展。2012 年 1 月，世界经济论坛年会把“大数据、大影响”作为重要议题。美国从开放政府数据、开展关键技术研究和推动大数据应用三方面布局大数据产业。美国在开放政府上非常积极，通过 Data.gov 开放 37 万个数据集，并开放网站的 API 和源代码，提供数千个数据应用。除了推动本国政府数据开放外，美国还倡导发起了全球开放政府数据运动，已有 41 个国家响应。美国政府还投资两亿美元促进大数据核心技术的研究和应用，把大数据与集成电路、互联网放在同等重要的位置，从国家层面推进。

2. 资本市场也对大数据钟爱有加

2012 年 4 月，大数据分析公司 Splunk 高调宣传大数据，引发投资者关注。12 月初，为企业市场提供 Hadoop 解决方案的创业公司 Cloudera 获得 6500 万美元融资，估值约为 7 亿美元。近期，高盛联席主席斯科特·斯坦福说：“投资大数据及其运用回报率最高”。大数据领域的企业并购热度也在上升，单笔平均并购金额方面，大数据超过云计算位居 IT 领域榜首，在总并购额上也位居第二。

3. 人才需求巨大

据盖特纳咨询公司预测，大数据将为全球带来大约 440 万个 IT 新岗位和上千万个非 IT 岗位。麦肯锡公司预测美国到 2018 年需要深度数据分析人才 44 万～49 万人，缺口 14 万～19 万人；需要既熟悉本单位需求又了解大数据技术与应用的管理者约 150 万，这方面的人才缺口更大。中国是人才大国，但能理解与应用大数据的创新人才目前还很稀缺。

4. 国内情况

大数据的火爆，也带动了国内学术界、产业界和政府对大数据的热情。2011 年以来，中国计算机学会、中国通信学会先后成立了大数据委员会，研究大数据中的科学与工程问题，科技部的《中国云科技发展“十二五”专项规划》和工信部的《物联网“十二五”发展规划》等都把大数据技术作为一项重点予以支持。其中工信部发布的物联网“十二五”规划上，把信息处理技术作为 4 项关键技术创新工程之一被提出来，其中包括了海量数据存储、数据挖掘、图像视频智能分析，这都是大数据的重要组成部分。而另外 3 项关键技术创新工程，包括信息感知技术、信息传输技术、信息安全技术，也都与“大数据”密切相关。

大数据的热潮触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。

1.4.5 展望大数据

1. 大数据推动信息产业创新

大数据是指一般的软件工具难以捕捉、管理和分析的大容量数据，一般以“太字节”为单位，大数据之“大”，并不仅仅在于“容量之大”，更大的意义在于：通过对海量数据的交换、整合和分析，发现新的知识，创造新的价值，带来“大知识”、“大科技”、“大利润”和“大发展”。信息管理专家涂子沛在其专著中如是定义大数据。

根据IDC(国际数据公司)的监测统计,2011年全球数据总量已经达到约1.8ZB,而这个数值还在以每两年翻一番的速度增长,预计到2020年全球将总共拥有35ZB的数据量,增长近20倍。

2. 大数据将改变经济社会管理面貌

大数据作为一种重要的战略资产,已经不同程度地渗透到每个行业领域和部门,其深度应用不仅有助于企业经营活动,还有利于推动国民经济发展。麦肯锡研究表明,在医疗、零售和制造业,大数据可以每年提高劳动生产率0.5~1个百分点。

大数据技术作为一种重要的信息技术,对于提高安全保障能力、应急能力,优化公共事业服务,提高社会管理水平,作用正在日益凸显。大数据技术还可增强安全保障能力。在国防、反恐、安全等领域应用大数据技术,能够对来自于多种渠道的信息快速进行自动分类、整理、分析和反馈,有效解决情报、监视和侦察系统不足等问题,提高国家安全保障能力。

3. 大数据存储管理挑战及管理技术

目前电信、金融、零售等行业希望通过大数据的分析手段来帮助自己做出理性的决策。特别是电信和金融行业表现得尤为突出,市场数据没有与用户消费数据打通。面临的第一个问题就是海量数据存储的问题。多数企业正在试图建设自己的数据中心来满足大规模数据量的产生,但是随着数据的进一步增多,很多数据的查询和分析性能急剧下降,有的数据中心甚至出现了无法响应的状况,为企业的业务带来了很大损失。

企业的CIO们有着这样的疑虑,什么样的数据管理策略才能够对数据进行有效的保护,而且在需要时能让数据变得有价值。只有数据与适合的存储系统相匹配,制定出管理数据的战略,才能高成本、高可靠、高效益地应对大量数据。对于企业来说,大数据首先需要解决的问题就是成本和时间效应问题。商机不容错过,存储数据管理可以通过自动化操作实现,备份和归档软件可让企业的关键数据分存在不同的区域,然后按照特定的业务需求,对数据进行提取、操作和分析,并形成企业所需要的目标数据。大数据面临的存储难题迎刃而解。

大数据的关注度在不断升温,而大数据管理的技术也层出不穷。在众多技术中,有6种数据管理技术普遍被关注,即分布式存储与计算、内存数据库技术、列式数据库技术、云数据库、NoSQL、移动数据库技术。其中分布式存储与计算受关注度最高。

分布式存储与计算架构可以让大量数据以一种可靠、高效、可伸缩的方式进行处理。因为以并行的方式工作,所以数据处理速度相对较快,且成本较低,Hadoop和NoSQL都属于分布式存储技术的范畴。

内存数据库技术可以作为单独的数据库使用,还能为应用程序提供即时的响应和高吞吐量,SAP的HANA是该技术的典型代表。

列式数据库的特点是可以更好地应对海量关系数据中列的查询,占用更少的存储空间,这也是构建数据仓库的理想架构之一。

云数据库可以不受任何部署环境的影响,随意地进行拓展,进而为客户提供适宜其需求的虚拟容量,并实现自助式资源调配和自助式使用计量。目前微软公司的SQL Server可以提供类似的服务。

NoSQL数据库适合于以下场景,即庞大的数据量、极端的查询量和模式演化。企业可

以通过 NoSQL 获得高可扩展性、高可用性、低成本、可预见的弹性和架构灵活性的优势，甲骨文在 2011 年推出 Oracle NoSQL 数据库。

移动数据库技术是适应移动计算的产物。随着智能移动终端的普及，人们对移动数据实时处理和管理的要求不断提高，移动数据库具有平台的移动性、频繁断接性、网络条件的多样性、网络通信的非对称性、系统的高伸缩性和低可靠性以及电源能力的有限性等，也正是因为这些特性被业界所重视。

4. 我国大数据发展策略

中科院计算所网络数据科学与工程研究中心主任程学旗在接受采访时表示："数据的规模如此之大，现有的 IT 技术根本没有办法分析处理，价值难以得到有效利用。对这些数据的感知、分析，同时加以商业化，就是大数据技术需要完成的工作。"如何挖掘大数据的价值是重中之重。

我国应将大数据作为新一轮科技竞争和产业竞争的战略重点和制高点，充分认识"数据、技术、应用"三位一体、有机统一的内涵，掌握未来大数据发展主动权。

第2章 Data Studio

本章将会介绍 Data Studio 相关操作内容，使用软件环境为 IBM DB2 10.5 版本。

2.1 创建数据库

2.1.1 命令方式创建数据库

本小节将会介绍使用命令方式创建数据库。

CREATE DATABASE 用于创建数据库。基本语法是：

```
CREATE DATABASE database_name;
```

当然，此命令还包括许多参数选项，这里不进行详细说明。

例如：

```
CREATE DATABASE MYDB AUTHMATIC STORAGE YES ON 'C:\' DBPATH ON 'C:\'
USING CODESET GBK TERRITORY CN COLLATE USING SYSTEM PAGE SIZE 4096;
```

此命令创建了一个名为 MYDB 的数据库。

创建一个名为 dbsample 的数据库，如图 2.1 所示。

2.1.2 通过数据库向导来创建数据库

创建数据库向导可以帮助用户创建新的数据库和调整现有的数据库。操作步骤如下。

(1) 打开 IBM Data Studio 并切换到【管理资源管理器视图】，如图 2.2、图 2.3 所示。

(2) 在【管理资源管理器】对话框中右击选择【DB2】|【新建数据库】命令，如图 2.4 所示。

(3) 输入相关认证信息，单击【完成】按钮进入数据库创建向导，如图 2.5 所示。

(4) 在【详细信息】对话框中，为新数据库指定信息，如图 2.6 所示。在【存储器】对话框中，指定存储数据的位置，如图 2.7 所示。