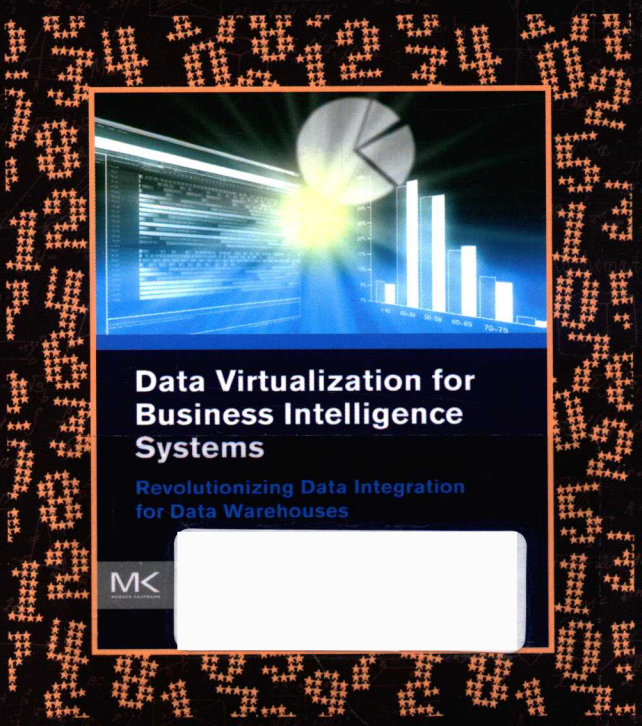
HZ BOOKS 华章IT
ELSEVIER 爱思唯尔

数据科学与工程技术丛书

数据虚拟化

商务智能系统的数据架构与管理

[荷] 里克 F. 范德兰斯（Rick F. van der Lans） 著
王志海 韩萌 孙艳歌 等译



DATA VIRTUALIZATION FOR BUSINESS INTELLIGENCE SYSTEMS

REVOLUTIONIZING DATA INTEGRATION FOR DATA WAREHOUSES

机械工业出版社
China Machine Press

数据科学与工程技术丛书

DATA VIRTUALIZATION FOR BUSINESS INTELLIGENCE SYSTEMS

REVOLUTIONIZING DATA INTEGRATION FOR DATA WAREHOUSES

数据虚拟化

商务智能系统的数据架构与管理

[荷] 里克 F. 范德兰斯（Rick F. van der Lans） 著

王志海 韩萌 孙艳歌 等译

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

数据虚拟化：商务智能系统的数据架构与管理 /（荷）里克 F. 范德兰斯著；王志海等译 .
—北京：机械工业出版社，2017.7
（数据科学与工程技术丛书）
书名原文：Data Virtualization for Business Intelligence Systems：Revolutionizing
Data Integration for Data Warehouses

ISBN 978-7-111-57612-9

I. 数… II. ①里… ②王… III. 数据处理 IV. TP274

中国版本图书馆 CIP 数据核字（2017）第 191499 号

本书版权登记号：图字：01-2013-1806

数据虚拟化：商务智能系统的数据架构与管理

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号　邮政编码：100037）
责任编辑：朱秀英　　责任校对：李秋荣
印　　刷：三河市宏图印务有限公司　　版　　次：2017 年 8 月第 1 版第 1 次印刷
开　　本：185mm × 260mm　1/16　　印　　张：14
书　　号：ISBN 978-7-111-57612-9　　定　　价：69.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88379426　88361066　　投稿热线：（010）88379604
购书热线：（010）68326294　88379649　68995259　　读者信箱：hzit@hzbook.com

序　言

经典数据仓库和商务智能的架构依赖于其中所存储的质量和集成程度都较高的数据本身。在商务智能的早期，我们努力手动地从多个操作型系统之中提取数据、组合数据、修复错误数据、填充缺损字段、删除重复数据，以及将最终集成的数据装载到当前数据库之中，以创建物理数据仓库或“单一数据源”，进而用于生成关于数据的汇总和分析报告。

之后，由于提取－转换－装载（ETL）工具的技术创新，从而能够以可靠的、可重复的方式来自动完成原本手动进行的数据集成任务。ETL 工具大大提高了创建数据仓库的整个过程的效率，其中包括数据质量技术，以进一步提高用于决策制定的集成数据的价值。到目前为止，ETL 工具仍然是商务智能系统中用于创建历史数据的物理存储的主要机制。

最近，操作型商务智能和大数据分析的出现使得商务智能架构师需要重新思考 ETL 和数据管理基础结构。首先对操作型商务智能来说，大多数商务智能化环境开始审视生成历史报告和分析的过程中发生了什么。历史数据也可以用于预测将会发生什么，但它并不能完全支持实时决策或者操作型商务智能。

随着企业开始逐渐需要某种能够基于当前或低延迟数据进行快速决策的能力，我们通过改变数据捕获技术、数据载入技术和操作型数据的微批量处理技术等来加速整个 ETL 过程。这些方法可以将数据仓库中数据的延迟从数天和数小时的级别降低到分钟，但它们仍然不足以实现真正的实时决策。商务智能实现者意识到经典的 ETL 处理已经达到极限，现在需要一种新的数据集成形式。

大数据及其相关的分析也面临着同样的问题。大数据包括社交和文本分析数据、传感器数据，以及事件或运动中的数据。许多大数据是非结构化的，或者更精确地说它有着多种格式。在典型的操作型系统中尚没有传统的、可预测的结构。相对于以前的标准，它的数据量极大。对于许多数据仓库实施者来说，他们面临着大数据整合的巨大挑战。

事实上，许多大数据并不需要永久存储在结构化的数据仓库中。通常，大数据需要经过实验性和调查性的分析。即使如此，也需要将一些数据与数据仓库中的数据进行组合。我们如何有效满足运营商务智能和大数据的需求，并扩展商务智能架构，而不会中断现有的 ETL 流程呢？答案就是数据虚拟化。

我认识本书作者 Rick F. van der Lans 多年。我经常学习他的文章和研究论文，因为他写的内容总是能对传统思维给出有趣的新思路并进行创造性的革新。Rick 的著作总是促使我重新评估现有的认知，而本书是他的最新著作。

数据虚拟化已成为当今商务智能实施者必备的技术。与其他新技术一样，它也面临

着许多问题。例如如何实施它，什么时候使用它，以及要避免什么陷阱。Rick 在这本实用的指南中详细介绍和解答了这些问题。之后几年我将会多次翻开这本书，我知道你也会。

Claudia Imhoff
Intelligent Solutions 公司总裁
Boulder BI Brain Trust（BBBT）公司创始人

前　言

概述

数据虚拟化是一种转化异构数据库集合和文件的技术，这种技术使得这些数据看起来像一种集成的数据。在用于商务智能系统时，它可以使数据架构更简单、更便宜，最重要的是更敏捷。新的报告和分析需求可以更快实施，现有系统可以更容易改变。这就需要增加敏捷性：一方面，商务用户需要其系统提供更多的敏捷性，因为他们的世界已经开始改变；另一方面，商务智能的新形式，如运营报告、大数据分析、360° 报告、自助服务报告和探索性分析，都是当前的需求。本书致力于数据虚拟化技术以及如何有效地在商务智能系统中利用该技术。因此让我们从起点开始，从虚拟化开始。

在 IT 行业，我们已经进入了虚拟化时代。似乎这一行业中的任何东西都可以虚拟化，包括内存、外存、网络和数据中心。虚拟化技术很热门，比如云技术的普及也可以归类为虚拟化技术。虚拟化技术就是热点，并且在一段时间内都将是技术的焦点。

所有虚拟化技术和概念的共同点是它们封装了某个资源。任何虚拟化解决方案都隐藏了可用资源的数量、资源的位置以及获取资源所需的 API 等。但不要将虚拟化与一些电脑游戏所提供的虚拟世界混淆。这些游戏提供了一些虚拟的东西，但它们并没有封装特定的资源。

本书将解释一种特定形式的虚拟化：数据虚拟化。简而言之，数据虚拟化意味着将数据以集成的方式提供给应用程序，而不管所有数据是否分布在多个数据库中，是否以不同格式存储，是否可通过不同的数据库语言进行访问。数据虚拟化技术将这些不同的数据存储作为一个逻辑数据库呈现给应用程序。虽然数据虚拟化产品和技术已经存在了一段时间，但是大约在 2009 年，它才逐步得到了应有的关注。因为它对解决方案的影响——增加的敏捷性，所以越来越多的组织正在采用虚拟化技术，似乎可以认为 21 世纪的第二个十年将成为数据虚拟化的十年。

数据虚拟化可以部署在需要检索和操作数据的所有类型的信息系统中，例如经典数据输入系统、基于因特网的系统、面向服务的系统、主数据管理系统和商务智能系统。本书的重点是商务智能系统。数据虚拟化可用于整合来自各种数据源的数据，包括数据仓库、数据集市和生产型数据库。它有潜力改变我们开发商务智能系统的方式。数据虚拟化将成为大多数这类系统的心脏。

简而言之，数据虚拟化允许我们使用更简单和更灵活的架构构建商务智能系统。如果你想知道怎么做和为什么，本书就是为你而写的！本书将详细描述数据虚拟化产品的工作原理、技术应用、应该做什么和不做什么，以及在商务智能系统中应用它的好处。

谁应该阅读这本书?

本书适合以下人群:

- 负责开发和管理数据仓库和商务智能环境的商务智能专家,以及那些想知道如何通过应用数据虚拟化技术来简化系统或实现更灵活的商务智能系统的人。
- 信息管理专家,想知道数据虚拟化对其职业的影响,以及数据虚拟化将如何影响信息管理、数据管理、数据库设计、数据清洗和数据分析等活动。
- 主数据管理专家,负责创建主数据管理系统,并希望了解如何从部署数据虚拟化中受益。
- 数据架构师,负责设计数据的整体系统架构,用来向特定组织的任何机构提供所需要的数据。
- 数据库管理员,必须了解数据虚拟化服务器的特性和限制,用于确定如何以及在何处可以有效且高效地应用此技术。
- 设计师、分析师和顾问,必须直接或间接处理数据虚拟化,以及想知道数据虚拟化所能做的和不能做的。
- 学生,想要学习数据虚拟化技术,以及理解数据虚拟化技术与其他数据处理相关技术的区别。

预备知识

关于数据仓库、商务智能和数据库技术的一般性知识是必需的。

术语和定义

遗憾的是,数据虚拟化和数据仓库领域中使用的所有术语并非都是明确定义的,这一点在本书中讲得很清楚。为了避免混淆,我们试图清晰地定义大多数术语。但是,我们不能保证本书中的定义与你的定义一致。

造成这种混乱的原因很多。第一个原因是,供应商纯粹为了区分自己的产品和竞争对手的产品而经常提出新的术语,但营销人员不定义术语,他们只使用这些术语并用一般术语来描述产品。在大家意识到这一点之前,我们都在使用那些定义不明确或根本没有定义的术语。第二个原因是,这个领域发展非常迅速,在较短的时间内就可能要为新思维创造某种术语,并提出权衡性定义。结果,我们可能会匆忙地选择一个术语,而经过仔细检查后才发现它并不合适。

写在最后……

对我来说,写一本书的感觉像是独自一人完成项目:坐在办公室里几个小时、几天、几个月,喝一杯茶,听最喜欢的音乐。但这不是一个独奏项目,一本书通常需要很多人合作而成,本书当然也是如此。因此,我要感谢很多人对我的帮助,感谢他们为本书做出的贡献,提出的想法和意见,以及对我的支持和耐心。

- 感谢 Jim Bean 和 Richard Hackathorn 的技术审查。他们的意见可能比他们意识到的更有价值。在我还在写作本书的时候就得到了他们的反馈意见,这使得整个项目相当鼓舞人心。要是我以前的所有图书都有像他们一样好的技术审校者该多好。

- 感谢 Claudia Imhoff 女士。她是各种商务智能书籍的作者和合作者，企业信息工厂的合作设计者，许多关于商务智能和相关主题的文章的作者，Boulder BI Brain Trust 公司的创始人，无数事件的发言人，感谢她为本书作序。由于她在商务智能领域的出色表现，因此她是做这项工作的最佳人选。我非常高兴当我向她提出请求时她毫不犹豫地答应了。我仍然感到荣幸。谢谢你，亲爱的 Claudia！
- 从开始写作的第一天起，我就得到了以下三个供应商的全力支持：Composite Software 公司、Denodo Technologies 公司和 Informatica 公司。特别感谢以下专家：Composite Software 公司的 David Besemer、Robert Eve、Kevin O'Brien、Ian Pestell 和 Jean-Philippe；Denodo Technologies 公司的 Suresh Chandrasekaran、Juan Lozano 和 Alberto Pan；Informatica 公司的 Diby Malakar、James Markarian、Bert Oosterhof、Ash Parikh 和 Lalitha Sundaramurthy。他们都专业且耐心地回答了我的技术问题。
- 就冲着他们愿意分享自己对数据虚拟化技术未来的看法，我也要感谢 Composite Software 公司、Denodo Technologies 公司和 Informatica 公司各自的 CTO：David Besemer、Alberto Pan 和 James Markarian。
- 这是我在 Morgan Kaufmann 出版的第一本书。现在本书已经在书店和互联网上开始销售，必须说这是一个明智的决定。与 Andrea Dierna 和 Robyn Day 合作是一种乐趣。他们通过这个项目给了我很多指导。多亏了他们，最终才有了这本可读性很强的书。这是一次重要的经验，我为我曾经无组织的写作过程而道歉。
- 在本书中，大多数例子都涉及一个示例数据库，它源于 Roland Bouman 和 Jos van Dongen 在他们的书《Pentaho Solutions：Business Intelligence and Data Warehousing with Pentaho and MySQL》中设计的数据库。我要感谢他们开发了这个数据库并允许我在本书中使用。特别感谢 Roland，感谢他审阅本书的部分内容，每个作者都应该邀请他做自己书稿的编辑。
- 感谢过去几年中从世界各地赶来参加我们数据虚拟化和数据交付平台研讨会的数以百计的技术人员。他们的意见和建议对本书是无价的。
- 多年来与独立分析师 Colin White 和好朋友 Mike Ferguson 在商务智能和数据虚拟化上的讨论给我编写本书带来了很大的帮助。我们已经认识了近 20 年，我一直尊重和高度评价他们对新技术的看法。
- 上面提到的所有人都对本书贡献颇丰，但有一个人对整个项目来说至关重要：我的“私人编辑”，也是我的妻子 Diane Cools。我们已经一起合作了 10 多本书，这些年来，很多大型项目都是我们一起完成的。事实上，如果没有她，我怀疑自己可能不会写一本关于数据虚拟化的书。一如既往，非常感谢亲爱的 Diane！

最后，我想请读者将有关本书内容的意见、观点、想法和建议都发送到电子邮箱 info@r20.nl。非常感谢大家的合作。我希望你阅读本书时能够获得的乐趣和我写本书时获得的乐趣一样多。

Rick F. van der Lans

荷兰，海牙

作者简介

Rick F. van der Lans 先生是一位独立的咨询师、演讲家和作家。他专注于数据仓库、商务智能、面向服务的架构和数据库技术等方面的研究。他是 R20/Consultancy 私人有限公司的总裁。Rick 一直在参与各种与商务智能、数据仓库、面向服务的架构、数据虚拟化和集成技术的应用相关的项目。

Rick van der Lans 先生是一名国际知名的演讲家。他有着在欧洲、中东、美国、南美和澳大利亚等多个国家和地区 20 多年的专业演讲经验。他经常应主流软件供应商之邀发表主题演讲。

他是诸多计算机相关书籍的作者，其中包括《 Myths on Computing 》。他的一些著作被译成多种语言，例如畅销书《 Introduction to SQL 》和《 SQL for MySQL Developers 》有英语、荷兰语、意大利语、汉语和德语版本，并在全球销售。

作为 BeyeNetwork.com 网站和白皮书的作者、欧洲数据仓库和商务智能年会的主席以及 IT 杂志专栏作家，Rick 与许多供应商保持着密切联系。

你可以通过以下方式与 Rick 联系：

Email：rick@r20.nl

Twitter：http://twitter.com/Rick_vanderlans

LinkedIn：http://www.linkedin.com/pub/rick-van-der-lans/9/207/223

目　录

第 1 章

数据虚拟化导论

1.1 引言

本章解释了数据虚拟化是如何应用于开发更灵活的商务智能系统的。通过应用数据虚拟化，系统将变得更容易被改变。我们可以开发新的报告，同时更容易、更快速地适应现有的报告。这种灵活性对商务智能系统的用户来说是一个重要的方面。他们的世界变化越来越快，因此他们所支持的商务智能系统必须时刻保持在同一节奏上。

首先，我们讨论发生在快速发展的商务智能行业上的变化。接下来，我们将讨论什么是数据虚拟化和应用这种技术到商务智能系统所带来的影响。为了更好地理解数据虚拟化，我们描述了它与其他技术和思想的关系，如抽象、封装、数据集成和企业信息集成。换句话说，本章给出了数据虚拟化的一个高级概述。这些主题都会在后续章节中详细讨论。

1.2 商务智能世界正在改变

机构开发商务智能系统的主要原因是为了支持和改善他们的决策过程。这就意味着这些系统的主要用户是那些决策者。例如，他们决定何时订购更多的原材料，如何简化采购流程，给哪些客户提供折扣，是采用其他公司进行运输还是自己内部运输，等等。

这个世界的决策制定正在改变，其中最大的改变是组织机构必须反应得更快。这就意味着必须更快地制定决策，但是可用来制定决策的时间越来越少（有时是至关重要的）。研究调查可以证明这种现象。例如，由阿伯丁集团在 2011 年 3 月所进行的研究表明，43% 的企业发现很难做出及时的决策（如图 1-1 所示）。管理者越来越觉得在某些业务事件发生后，他们用来做决策的时间更少，结果是他们不得不更快地改变现有报告，并更快地开发新的报告。

但是这并非是唯一的改变。现在新的数据资源可以用于分析和报告，而在商务智能开始的那些年，对于报告和分析来说唯一可用的数据就是和业务流程相关的内部数据。例如，建立系统用来存储所有的订单、客户数据和发票。现在，越来越多的系统能够提供有价值的数据，如博客、电子邮件服务器、呼叫中心系统和文件管理系统。分析这类数据可以更好地理解客户怎么看待一家公司的服务和产品、网页具有怎样的高效性以及找到“好”客户的最好办法。

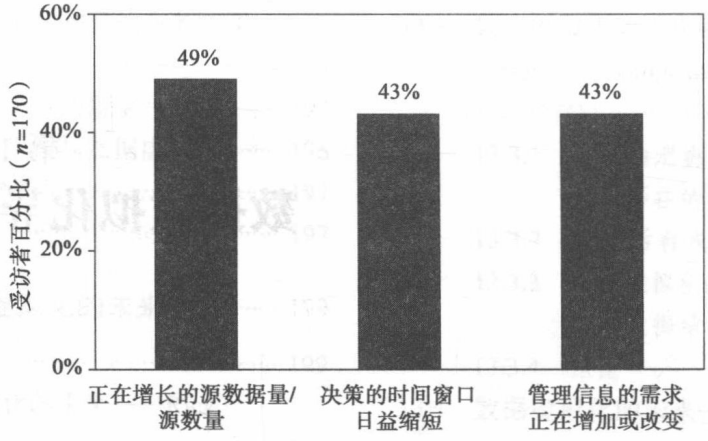


图 1-1　阿伯丁集团的研究。中间的柱状统计表明 43% 的受访者表示决策的时间窗口正在缩短，见文献［2］

也可以在某一机构或组织之外找到新的数据资源，例如，网页、社交媒体网络和政府数据都可能含有一些数据，这些数据与内部数据结合后会带来新的见解。组织机构喜欢将他们自己的内部数据和那些最新数据源结合，从而增强分析和报告能力。

在各个领域——无论是医疗、医药、电信、电子领域，还是在商务智能情况下——新技术创造了新机会，也因此正在改变这些领域。对商务智能来说，新技术变得可用，包括分析型数据库服务、移动商务智能工具、数据库内分析、大规模内存、高度并行的硬件平台、云和固态磁盘技术。所有的这些新技术都会明显扩展组织机构的分析和报告能力：它将支持各种形式的决策，这些决策大多数组织机构甚至还没考虑过；并且它将允许组织机构在几分钟时间内完成用旧技术花费几天时间才能完成的数据分析。

另一个明显的改变是关于对应用商务智能感兴趣的新用户群组的。目前，大多数商务智能系统的用户是那些处于战略和战术管理层次的决策者。在大多数情况下，这些用户可以完美地利用并不是百分百最新的数据。哪怕一天、一周，甚至可能是一个月的数据，对他们来说也是绰绰有余的。发生改变的是目前工作在操作层次上的决策者们被商务智能所吸引。他们理解它的潜在价值，因此想利用报告和分析工具的力量。然而，在大多数情况下，他们不能操作旧数据。他们想要分析百分百最新（或者至少接近百分百）的操作型数据。

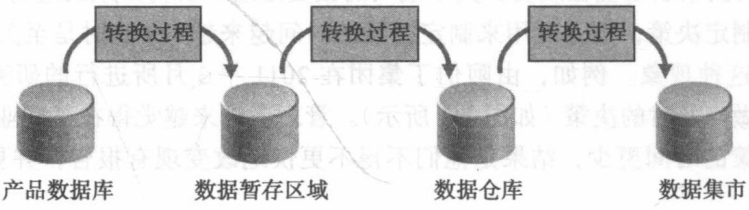


图 1-2　很多商务智能系统基于一个数据库和转换过程链

所有这些改变，特别是更快的决策制定，很难在目前的商务智能系统中实现，因为这些系统要求彻底的重新设计。这是因为大多数在过去 20 年开发的商务智能系统都基于一个数据库链（如图 1-2 所示）。数据从一个数据库转换和拷贝到另一个数据库，直至它到达一个终点：一个由报告或分析工具访问的数据库。每个转换过程提取、清洗、集成和转换数据，然后将它下载到数据库链上的另一个数据库中。这个过程持续到数据达到一个适用于报告和分

析工具的高质量水平和形式。这些转换过程通常称为 ETL（提取 – 转换 – 装载）。

这条数据库链和转换过程是长的、复杂的并且高度内部关联的。报告或数据的每一个改变都可能会导致整条链上无数的改变。在整条链上实现一个明显的、简单的改变可能要花费几天，甚至是几周时间。影响就是商务智能部门不能跟上业务要求改变的速度，这将导致应用积压并在决策速度和组织质量上带来消极影响。另外，由于需要很多的转换过程并且每个过程都需花费时间，所以很难在数据库链的终点交付操作型数据，如数据集市。

我们需要的是一个灵敏的、易于改变的体系结构，并且最好的办法就是创造一个包含更少组件的体系结构，这就意味着包含更少的数据库和更少的转换过程。少量的组件意味着需要改变的东西将更少。此外，更少的组件简化了体系结构，也提高了灵活性。

这就是数据虚拟化的由来。简而言之，数据虚拟化是一种将可用数据转换成分析和报告所需形式的可选择技术。它需要更少的数据库和更少的转换过程。换句话说，在商务智能系统中使用数据虚拟化会得到更短的数据库链，需要开发和管理更少的数据库，并且将会有更少的转换过程。总之，应用数据虚拟化简化了商务智能体系结构，并因此形成更灵活的商务智能体系结构，这一结构符合当前组织机构的商务智能需求：简单就是更灵活。

1.3　虚拟化简介

*数据虚拟化*这一术语是基于*虚拟化*的。虚拟化在 IT 产业并不是一个新概念。虚拟化的第一个应用大概是在 20 世纪 60 年代，IBM 使用这一概念将主机分成独立的虚拟机，这使得一台机器可以并行地运行多个应用程序。还是在 20 世纪 60 年代，虚拟存储通过使用一种称为*分页*的技术引入。内存虚拟化用来模拟比一台机器可用物理内存更多的存储空间。现在，几乎所有东西都能虚拟化，包括处理器、存储（见文献 [3]）、网络、数据中心（见文献 [4]）和操作系统。虚拟机和云也可以被视作虚拟化技术。

一般来说，虚拟化意味着应用程序可以使用一种资源而不需考虑它存储在什么地方，技术接口是什么，是如何实现的，使用的平台是什么，以及有多少是可用的。虚拟化方法封装资源使得所有的技术细节都隐藏起来，并且应用程序可以使用一个更简单的接口进行工作。

我第一次参加资源虚拟化的项目是在我职业生涯的早期。该项目需要写一个应用程序使得用户可以使用不同的接口技术进行工作。一个叫作*图文电视*，是为电视设备开发的；另一个叫作*基于字符的终端*。作为技术设计者，我决定开发一个 API，使得应用程序可以使用它从屏幕上获取数据并且返回输入。这个 API 是一个软件层，在其他应用程序使用时它会隐藏用户接口技术。在不了解它的情况下，我设计了一个用户接口虚拟化层。然后，我总是试着使用这样的方法来设计系统，为了简化开发，这一方法的某项特定技术的实现细节对于应用程序的其他部分是隐藏的。

1.4　什么是数据虚拟化

数据虚拟化是虚拟化的一种形式。正如这一术语表明的，它封装的资源是数据。简而言之，当应用数据虚拟化时，它提供了一个中间层，这个中间层对应用隐藏了大多数关于数据是怎样存储、存在哪里这些方面的技术部分（如图 1-3 所示）。因为这一层，应用不需要知道所有数据在物理上的存储位置，数据库服务器的运行位置，需要的 API 是什么，使用哪种数

据库语言，等等。对于每个使用数据虚拟化的应用来说，它感觉像是在访问一个大数据库。

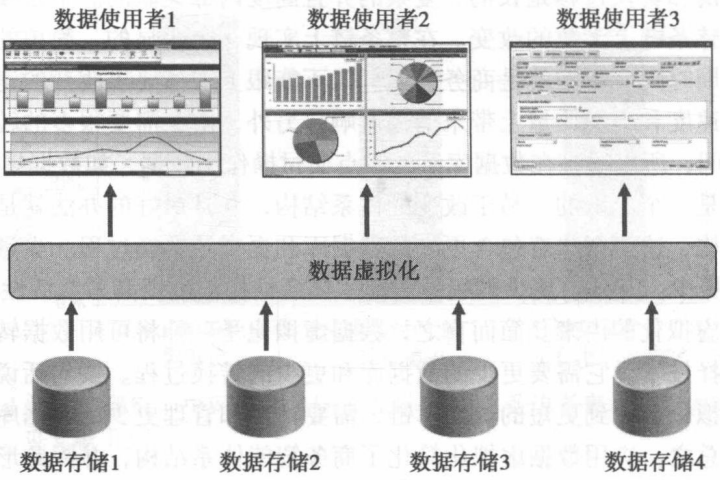


图 1-3　当应用数据虚拟化时，所有的数据资源都将以一个集成的数据源呈现

如果我们使用前几节中对虚拟化的描述来描述数据虚拟化，我们认识到：数据虚拟化意味着应用程序可以使用数据而不用考虑它存储在什么地方，技术接口是什么，是如何实现的，使用的平台是什么，以及有多少是可用的。数据虚拟化方法封装数据资源使得所有的技术细节都隐藏起来，并且应用程序可以使用一个更简单的接口进行工作。

在图 1-3 和整本书中，我们使用*数据使用者*和*数据存储*这些术语。术语*数据使用者*指任何检索、输入或操作数据的应用程序。例如，数据使用者可能是在线数据输入程序、报告程序、统计模型、网络应用、批处理应用或 RFID 传感器。同样，术语*数据存储*用来指代任何数据源：SQL 数据库中的表、简单的文本文件、XML 文档、电子表格、Web 服务、顺序文件、HTML 页面，等等。在某些情况下，数据存储文件仅仅是某种被动访问的文件；而其他情况下，数据存储系统包含访问其本身的软件，例如数据库服务器和 Web 服务。

数据虚拟化方法存在于数据使用者和数据存储之间。数据使用者通过数据虚拟化层访问数据，数据虚拟化层隐藏数据存储。

1.5　数据虚拟化与相关概念

数据虚拟化和 IT 产业中一些众所周知的其他概念密切相关，如封装、信息隐藏、抽象数据联合、数据集成和企业信息集成。本节解释这些概念和它们与数据虚拟化的关系。

注意，关于这些概念存在不同的定义，并且有些人视这些概念为同义词。混淆这些概念的原因在文献［5］中 Edward Berard 已经表述得相当清晰："抽象、信息隐藏和封装是非常不同但高度相关的概念。不难看出抽象、信息隐藏和封装是怎样相互混淆的。"

1.5.1　数据虚拟化与封装和信息隐藏

40 多年前，David L. Parnas 在 1972 年发表了开创性论文"On the Criteria to Be Used in Decomposing Systems into Modules"（见文献［6］）。该论文已经出版了多次。在这个传奇的论文中，Parnas 解释了对于应用程序来说，开发中将应用与存储的数据结构独立开是多么重