

数据库技术丛书

数据建模

分析与设计的工具和技术

Data Modeler's Workbench

Tools and Techniques for Analysis and Design



(美) Steve Hoberman 著

贾爱霞 程耀 译



机械工业出版社
China Machine Press

数

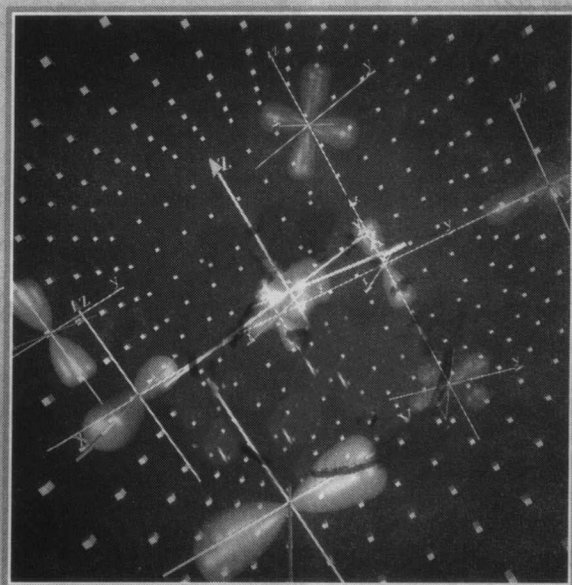
书

数据建模

分析与设计的工具和技术

Data Modeler's Workbench

Tools and Techniques for Analysis and Design



(美) Steve Hoberman 著

贾爱霞 程耀 译



机械工业出版社
China Machine Press

本书包含20多个经过精心调试的工具,这些工具的目的在于改进数据库、数据仓库和操作型应用程序的速度、准确性、适应性和一致性。Steve Hoberman通过示例阐释了每一个工具,展示了如何应用每个工具,以及各工具在操作型和发布型环境中的什么地方才是最有效的。你可以为特定的行业、组织或项目定制本书中的工具。

Steve Hoberman: Data Modeler's Workbench: Tools and Techniques for Analysis and Design(0-471-11176-9).

Authorized translation from the English language edition published by John Wiley & Sons, Inc.

Copyright © 2002 by Steve Hoberman.

All rights reserved.

本书中文简体字版由约翰·威利父子公司授权机械工业出版社独家出版。未经出版者书面许可,不得以任何方式复制或抄袭本书内容。

版权所有,侵权必究。

本书版权登记号:图字:01-2003-2883

图书在版编目(CIP)数据

数据建模:分析与设计的工具和技术/(美)霍伯曼(Hoberman, S)著;贾爱霞,程耀译.-北京:机械工业出版社,2004.6

(数据库技术丛书)

书名原文:Data Modeler's Workbench: Tools and Techniques for Analysis and Design
ISBN 7-111-14308-6

I. 数… II. ①霍… ②贾… ③程… III. 数据库技术-建立模型 IV. TP311.13

中国版本图书馆CIP数据核字(2004)第028918号

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:李云静

北京中兴印刷有限公司印刷·新华书店北京发行所发行

2004年6月第1版第1次印刷

787mm×1092mm 1/16·24.5印张

印数:0 001-4 000册

定价:49.00元

凡购本书,如有倒页、脱页、缺页,由本社发行部调换
本社购书热线:(010) 68326294

译者序

Steve Hoberman的这本书很值得一读。

所谓工具和技术，就是作者自己在建模工作中创建和使用的一些电子表格，它们就好像放在木工、瓦工工作台上的锤子、钳子和螺丝刀一样，它们就是数据建模者工作台上的各种必不可少（作者的观点）的工具。

目前国内的软件设计人员可能很少会把这些表格看作工具。但是，通过研读这本书，我们认为，这些表格确实能够帮助大家自上而下、由整体到局部地分析和验证数据需求、寻找数据中的缺陷和矛盾、建立正确的规范化和反向规范化的途径，并最终形成逻辑完善、布局美观而实用的逻辑和物理数据模型。

我们还感到，国内的系统分析和软件设计人员可能不太善于总结，他们很少把自己的多年经验归纳、分析，形成自己的分析和设计风格，更不用说形成自己的一套工具了。毋庸置疑，经常对自己的工作进行总结有利于使自己的水平有一个质的提升，也可以给周围的同行提供实用的经验。总之，在我们应付手下的繁重工作的同时，还有必要争取做一名“思想者”。

与以往的数据库技术的书籍不同，本书还突出了数据建模这个过程——不仅是一个技术过程，还是一个人文的过程。因为决定一个项目成败的不仅仅是技术问题，还有很多其他因素。尽管他们的文化与我们存在着不小的差异，但本书仍能给我们很多启迪。

参加本书翻译工作的还有：贾森、张宏、叶明、李斌、王小捷、李娜、吴示成、赵孟文。他们在维护术语列表、正文核对、录入和打印文稿方面作出了贡献。

序

如果把专业建模者在日常工作中所面对的所有挑战画成一幅地图，然后再配上文字解说，来评价在每一个挑战面前他们的表现，那将很有趣。这将是一幅很不规则的图片。

规范化可能会赢得“最受关注话题”奖。的确，一个局外人可能会轻易地得出结论，数据建模主要就是指复杂规范化问题的解决方案，数据建模者的主要工具就是有关的代数和算术。

使用语言和示意图表示法的惯例受到了广泛的追捧，我不否认它的重要性，但它们是否真的如此重要？这些年来，不断出现其他的供选择的方法，以及关于这些方法的适用性的许多争论——有时甚至是激烈的争论，这已经形成了一种稳定的潮流。

现在由于近期的一些出版物（这些出版物大都出自专业人士），才使我们在从前的数据模型模式和业务规则的这片“贫瘠的土地”上也看到了某些有益的著作。

以上这些内容都很重要，但它们只针对数据建模者实际所做工作的一部分。惯例的选择通常都是一次性的决策，一般由其他的某个人来做出。如果最终规范化被完成了，它通常是对已经凭直觉完成的事情的一个最终检查。自从前三层范式的原始阐述出现之后，30年的研究对在实际中生成的模型并没有什么影响。

这种失衡的状况所缺少的就是关于数据建模过程方面的材料：项目计划的实用性，与最终用户的沟通，验证规则和假设，向不同的听众讲述模型，甚至说服股东无论如何数据建模都是必要的。这些都是杂乱无章的、达不到科学意义上的精确性的素材，它们消耗真正的数据建模者的时间，但常常能决定着一个项目是成功还是失败。

直觉给出的答案是，这些东西我们可以从实践中得到，无须从书本中学习。这不应该是一个可接受的答案。的确，许多这样的主题对研究者来说可能缺乏吸引力：它们不容易分类，它们需要实践和组织设置方面的具体知识，而且很少有单一的解决方案。但是，如果我们要避免重新摸索规律，或者避免项目失败，那么就存在整理经验并系统化的必要性。

Steve Hoberman在本书中触及了许多人们过去尚未涉足的领域。他描述了数据建模者所面对的一系列挑战，这本身就是一个成就。其中的许多挑战都是我们所有人所面对，但从来就没有静下心来进行思考或讨论的共同问题。

然后，他提出了自己的解决方法：工具、思想、沟通的方式。它们是正确的，人们认识到，它们的出现是必然的，它们在反复的应用中逐渐发展。对于新手来说，这些方法会提供一个极好的出发点；对于行家来说，它们会引发一次对工具的重新审视，并且促使他们更新自己的工具集合。

Steve并没有在本书中提供一个方法论或秘诀。但数据建模者却会在书中找到“一付药方，包治百病”的解决方案。我希望，这个行业成熟到这样一种程度：我们认识到，我们所需要的就是一个广博的公用知识体系，以及适应这个体系的能力。

在本书中，我们不仅能看到对数据建模行业的知识体系的实质性的贡献，而且还能看到一个引起争论和探讨的新领域的概貌。我希望，从业者和研究者都要勇敢地面对这个挑战，来进一步探索关于数据建模实践的这些重要领域。

Graeme Simsion

2001年9月

前 言

你是否知道自己在准备外出度假之前，总是往行囊里装哪几样东西？我发现，我自己常装的是牙刷、旅行闹钟、剃刀、剃须膏等等。这些东西对我来说十分重要，它们能保证让我度过一个舒适的假期。如果我忘记带其中的某些东西，我通常不得不在当地的商店里再去购买。如果事先把它们都带齐，我就能节省再次购买所花的时间和金钱。

同样，当我们准备踏上数据建模之旅时，也有一系列东西或者工具，我们需要携带着它们，它们会在设计时派上用场。这些工具以电子表格、指南、问题、提示等形式存在，使用它们会使数据建模工作进行得更有效率，并最终形成一个高质量的设计。使用已有的数据建模工具，较之根本不使用工具或创建你自己的工具，要节省时间和开销。本书中包含了一组工具，它们使得数据建模过程进行得更有效率，并可产生高质量的数据模型工作产品。

本书及技术概览

数据模型是应用程序的核心和“灵魂”。一个好的数据模型是高效的数据输入和获取的基础。同时，它还需要与其他模型保持一致，至少要与你的组织中的其他模型一致，并且要能准确地记录恰当的业务需求。一个好的数据模型必须使应用程序能适度地扩展和进化，以便支持不断变化的业务需求。尽管数据模型有如此高的价值，但我还是很诧异：我们对于创建数据模型的过程所给予的关注常常如此之少，并且我们也很少依靠工具和技术来不断改善数据建模的过程。

当我在教授数据建模时，我通常从一个小练习开始，在这个练习中，我要求学员们根据一个特殊的表格或报告来创建一个数据模型。当我在教室中四处巡视时，我能看到人们以很多不同的方式创建模型，这种情形总是给我留下深刻的印象。有些人直接得出一个数据模型，有些人则从电子表格开始，更有一些人列出了所有的数据元素。他们中有些人能在规定的10分钟内正确地完成这个小小的课堂建模练习。另外一些人则可能花费了10分钟却几乎尚未触及问题的皮毛。但有一点是可以肯定的：那些正确地完成了建模的人使用了数据建模过程或工具。

从1990年开始，我就一直从事数据建模工作，并开发了一套我自己使用的过程和工具。本书包含了一组经过精心调试的数据建模工具，而且这些工具是按照它们在建模生命周期中应该出现的次序来组织的。这些工具将帮助我们捕捉需求、为需求建模并确认数据需求，以改进数据仓库和操作型应用程序的速度、准确性、适应性和一致性。

本书有两个主要目的：

- 1) 提供一组目前就可以在你的数据模型设计中开始使用的数据建模工具。本书将带你亲历几个数据建模工具，我们从几个在进行分析和建模之前就需要应用的工具开始，最后以“在模型中适当地摆放数据元素”结束。

- 2) 激发你的想像力来创建你自己的工具或改造本书中的工具。我不可能讨论所有的建模工具，在你读完本书之后，你应该能够设计出对你的课题、组织和行业最适合的、你自己的工具。

我之所以写作本书，是因为我相信，使用这些已被证明能改进设计的数据建模工具，可以大量节省数据建模者的时间、精力，并且避免他们创建自己的数据建模工具时的苦恼和错误。本书填补了我们这个行业中的一个空白，因为现有的许多关于数据建模的书所讨论的都是“如何进行”形式化的数据建模，或者是针对某种类型的应用程序，如数据仓库。目前所缺少的正是一本关于改进数据建模过程和数据建模产品的实用工具方面的书。这些实用工具能被应用到任何一类应用程序；无论它们本质上是发布型应用程序还是操作型应用程序，无论你所处的行业或使用的应用程序方法论是怎样的，这些工具都适用。

我知道，在你阅读本书的过程中，在某些地方，你可能会自问：“这些技术是数据仓库设计所用的数据建模技术，还是操作型应用程序设计所用的数据建模技术？”答案是，本书中绝大多数的技术适用于所有类型的应用程序。针对发布型应用程序和操作型应用程序，这些工具中的某些工具的答案或结果可能会不同；然而，这些工具对这两种应用程序都适用。这个一般规律也有例外情况：在本书中讨论特定的工具时，我会说明这一点。

本书的组织方式

按照这些工具在数据建模过程中应该被使用的时间，本书分为三部分。例如，第5章中的“主体域家族树”(Subject Area Family Tree)应该总出现在第7章中的“数据元素家族树”(Data Element Family Tree)之前。尽管我建议按顺序阅读本书正文，但当你读完了第2章和第4章之后，你将对本书的其余部分的顺序有一个很好的理解；这样，你可以直接跳到那些你最感兴趣的章节。但在每一章中，我都会提到以前的工具，以便使本书自始至终保持逻辑上的连贯，方便阅读。

下面是这三个部分的概览：

- **第一部分：建立基础。**第1章到第4章提供了几个工具，以帮助数据建模者进行数据建模训练和交流。
- **第二部分：分析需求。**第5章到第7章主要关注一个应用程序的主体域需求的记录和确认，还关注将为我们的逻辑数据建模阶段提供输入的数据元素细节。
- **第三部分：建立需求模型及相关建议。**第8章到第10章关注如何为我们的应用程序创建最佳的逻辑和物理模型。我将提供一系列工具，它们的目的是以一种灵活的方式来建立数据需求的模型，同时还考虑数据库的性能和存储的结果。第11章是奉送给大家的“赠品”，它也被包括在第三部分。第11章是本书的结尾，它给你一些忠告和建议，这是我在多年的经历中学到的东西，而且作为一名数据建模者，我一直遵循这些原则进行建模工作。

下面是对每一章的小结：

第1章：“使用趣闻、类比和演示文稿来阐明数据建模的概念”。该章关注就数据建模的各个方面进行沟通的最有效的方式。趣闻就是小故事，它们对阐明一个概念和术语非常有效。类比把数据建模概念与听众非常熟悉的某个事物相比较，以强调某些数据建模特性。演示文稿制作步骤就是创建一个数据建模演示文稿时要经历的一系列有序的步骤。

第2章：“元数据宾果游戏”。该章帮助你的项目和组织确定标准类型的元数据。“元数据宾果游戏”是一个游戏，在这个游戏中，你的项目团队中的人们完成宾果卡片，从而确定哪些类

型的元数据最重要、最需要记录。这些宾果卡片记录处于不同层次的元数据，包括项目层次、数据模型层次、主体域层次、实体层次和数据元素层次的元数据。

第3章：“确保高质量的定义”。该章关注一个称为“定义检查单”的工具。该章还提供了在主体域、实体和数据元素层次应用“定义检查单”的例子。

第4章：“数据建模者的项目计划”。该章包括四个项目计划模板，这些项目计划模板包含一组完整而连续的任务，这些任务是在一个现实的时限范围内成功地完成数据建模工作产品所必需的。

第5章：“主体域分析”。该章为确定和记录一个应用程序的主体域提供了完成相关工作产品的工具。“主体域检查单”是新应用程序中的主体域及其定义的一个完整列表。“应用程序主体域CRUD（创建、读取、更新和删除）矩阵”包含存在于你的新应用程序和已有的应用程序之间的主体域的差异和重叠之处。这是发挥你的应用程序能力的一个有力工具。“In-the-Know模板”确定为完成这个新应用程序的数据建模工作产品所需要的资源，这些资源包括人员和文档。“主体域家族树”包含每个主体域的源应用程序和一些其他的关键信息。“主体域粒度矩阵”以一个电子表格的格式来记录每个度量或事实主体域的发布层次。

第6章：“主体域建模”。该章关注三种类型的主体域模型，每一个模型都是一个非常有力的确认和检查工具。“业务清理板”模型帮助我们理解独立于任何应用程序的业务领域。“应用程序清理板”则以“业务清理板”为基础，只关注对应用程序来说至关重要的东西。“早期现实性检查”模型把新的应用程序与一个已有的应用程序体系结构相比较，以便确定它们之间的重叠之处及其影响。

第7章：“逻辑数据分析”。该章呈现了四种有效的数据元素记录和确认工具。“数据元素家族树”记录数据元素名称、别名、定义、业务目标、缺省值、变换关系、格式和可置空特性。“数据元素粒度矩阵”记录事实和它们的发布层次之间的关系。“数据质量记录模板”包含数据和元数据之间的准则和对照信息，“数据质量确认模板”则提供这样的保证：每个数据元素的数据质量都已经经过适当地检查。

第8章：“规范化之旅和反向规范化生存指南”。该章给出了对你的数据需求进行规范化和反向规范化的工具。“规范化之旅”就是一系列规则和指南，它们能够引导你达到全面理解应用程序的顶点。“反向规范化生存指南”就是一个问答式的方案，它用于确定在何处对逻辑数据模型进行反向规范化。当你对每一个关系都完成了这些问题的答案时，你就有了一个处于适当的反向规范化层次的物理数据模型。

第9章：“抽象化安全指南与组件”。该章帮助我们最大程度地减少未来的数据需求对模型和最终的数据库设计的影响。“抽象化安全指南”是一组三个问题，它们将帮助你确定你必须对数据模型的哪些地方进行抽象化。“抽象组件”工具是那些最有用的抽象构建块的集合。我在这一章中所讨论的这些组件，你可以用在你自己的应用程序模型中，用于替代你的模型中需要更强适应性的部分。这些组件可以存在于实体、关系或数据元素的层次。

第10章：“数据模型美化技巧”。使你的设计超越直接的应用程序需求，关注改进逻辑和物理数据模型的视觉外观的技巧。这些技巧是在实体、关系和数据元素等细节层次的基础上提供的。

第11章：“规划一个长盛不衰的数据建模生涯”，主要讨论我的一些忠告；作为一个数据建模者，我也遵从这些忠告。这些忠告包含了我多年来学到的东西；有些忠告来自于我自己的体会，有些忠告来自于我周围人们的成功的经验和失败的教训。遵从这些忠告将有助于你成为一个更成功的数据建模者。这些忠告是以“十大……”清单的形式表达的。

本书中的表示法和结构

在本书的各个章节中，我在显示信息时使用了几种统一的格式。

注意 在本书中，我是按照这种格式来显示“注意”、“警告”和“提示”的。这样，在你阅读每一章时，就可以很容易地辨认出这些信息。“注意”对所讨论的话题提供额外的信息，“警告”包含你应该避免的缺陷和问题，“提示”则是简短而重要的忠告话语。

我使用楷体来描述你可能在某一天发现自己会同样面对的某种现实生活情形。例如，在第7章中，你会看到如下的故事：

你已经完成了应用程序的所有主体域分析工具，包括“主体域家族树”。现在该做什么了？你必须确定每个主体域中的数据元素，以及它们的来源和可能的变换。你必须把“主体域家族树”带领到下一个层次。在这个细节的层次上，你要确定的是将在以后的开发过程中使用的数据元素。这意味着，我们在这个阶段没有识别出的数据元素可能就不会出现在最终的应用程序中。因此，极重要的是，尽可能以一种简洁而清晰的方式来组织这些数据元素，以便最大程度地记录最完整、最准确的信息。我们需要“数据元素家族树”。

另外，在本“前言”和每一章的开头部分，都有一个小故事，它把一个个人的休假经历与每一章中的工具联系起来。在我的内心世界里，我发现有两个领域对我来说非常亲密和亲切，这就是数据建模和休假。所以，为什么不把它们两个结合到每一章中呢？

你还会注意到，每一章中都有如下类似的章节：

本章概览。用几个段落来描述本章中包含的内容、将被讨论的话题的顺序，以及本章中每一个工具的一个非常简短的介绍。该节中通常还有几句话，用于说明本章和本书中前面的章节有何联系。

该工具是什么？关于该工具的解释，包括它的目的，通常还有几个简单的例子。注意，对于每一章中的每一个工具，该节和下面两小节都会重复介绍。

该工具的使用方法。包括你使用该工具时要采取的步骤。

该工具的使用演练。包含按照“该工具的使用方法”一节中所描述的步骤来应用该工具的一个详细的例子。

小结。简要地总结该章的要点，并把该章与后续的章节相联系。

什么人应该阅读本书

如果你对数据建模和数据建模概念有一定程度的了解，那么，本书就是为你准备的。适合

阅读本书的人包括数据建模者和管理者、分析员、设计者以及系统和数据设计人员。如果数据建模对你来说是完全崭新的课题,那么,你就不宜阅读本书;因为,有许多关于数据建模的很好的入门性的教科书会更适合你。本书结尾处的推荐读物清单中列出了一些较好的入门性教科书。如果你已经有了一些实际的数据建模实践,就应该阅读本书。我坚定地认为,即使已有了10多年的数据建模实践,也可能会在本书的字里行间找到几个有用的工具和技术。

如果你有3年以下的数据建模实践,那么,我强烈地建议你阅读本书中的所有内容。如果你有3年以上的数据建模实践,那么,我建议你从第4章开始阅读,该章给第5章到第10章提供了一个很好的概述,而且会帮助你直接定位到本书中最吸引你的章节。如果你有了许多实践经验,那么,我建议你跳过第8章中的“规范化之旅”这一节,而只阅读第8章的后一部分,即“反向规范化生存指南”。对于所有的读者,我都建议大家在这个时候阅读第11章。尽管这一章很短,但它包含了对于数据建模者的有价值的忠告。

我还要建议你在本书中标记出你最感兴趣的那些章节,并在你进行自己的数据建模活动时回头参考这些章节。许多时候,当我们阅读一本书时,在读完每一章时,我们都点头表示理解,但是,在实际进行一项工作时,我们却忘记了应用所掌握的信息。我建议,你在进行数据建模时把本书作为一个有效的参考书来使用。例如,你可能会问自己:“我该怎样组织‘数据元素家族树’?”,然后迅速在第7章中寻找答案。

Web站点有哪些信息

如果你访问我的Web站点www.wiley.com/compbooks/hoberman,就会发现另外的工具和模板,以及本书中已有工具的更新版本。如果你不想重新制作本书中的某些电子表格,就可以访问这个站点,从中下载一个空白模板,并把它作为你的建模起点。我还将继续把新的工具和对于数据建模者的忠告更新到这个站点中。这个站点还包含一些链接,它们指向一些我最喜欢的数据建模Web站点。

你可以通过与本书配套的Web站点或者me@stevhoberman.com与我联络。

从这里开始

我是基于如下的意图来写作本书的:让这本书非常实用,而且让它包含容易应用到任何类型的应用程序的信息。我希望你在阅读这些工具的有关内容时发挥你的想像力,创造出新的工具。设想出新的工具或者对本书中的工具进行改造都会对你目前的任务有直接的影响。

随着数据建模的重要性日益提高,对数据建模者的要求将会更高。当我们以更加积极的眼光来看待数据建模时,就会发现,对数据建模者会有更多的挑战和更令人兴奋的机会。这些机会就是更迅速的数据模型转换和更灵活、更有效的设计结果。换句话说,在我们目前的环境中,甚至在不远的将来,诸如本书中的那些工具将成为一个成功设计的关键。

目 录

译者序
序
前言

第一部分 建立基础

第1章 使用趣闻、类比和演示文稿来

阐明数据建模的概念2

1.1 本章概览2

1.2 什么是数据建模趣闻3

1.2.1 数据建模趣闻的使用方法4

1.2.2 数据建模趣闻的实际演练5

1.3 什么是数据模型类比7

1.3.1 数据建模类比的使用方法8

1.3.2 数据建模类比的实际演练8

1.4 制作演示文稿的步骤是什么22

1.4.1 制作演示文稿的步骤24

1.4.2 实践创建演示文稿的步骤25

1.5 小结27

第2章 元数据宾果游戏28

2.1 本章概览28

2.2 什么是“元数据宾果游戏”29

2.2.1 理解元-元数据33

2.2.2 谁来参加“元数据宾果游戏”36

2.2.3 如何进行“元数据宾果游戏”38

2.2.4 “元数据宾果游戏”的评分过程54

2.3 “元数据宾果游戏”的实际演练54

2.4 小结56

第3章 确保高质量的定义58

3.1 本章概览58

3.2 什么是定义58

3.3 什么是“定义检查单”59

3.4 “定义检查单”的使用方法60

3.4.1 清晰61

3.4.2 完整67

3.4.3 准确75

3.4.4 文法77

3.4.5 长度77

3.4.6 “定义检查单”的实际演练78

3.5 小结87

第4章 数据建模者的项目计划88

4.1 本章概览88

4.2 什么是“数据建模阶段的工具”91

4.3 什么是“阶段-任务-工具”95

4.3.1 “阶段-任务-工具”的使用96

4.3.2 项目策划97

4.3.3 主体域分析98

4.3.4 主体域建模103

4.3.5 逻辑数据分析106

4.3.6 逻辑数据建模109

4.3.7 物理数据建模114

4.4 什么是“优先级三角形”115

4.5 什么是“可靠的估算工具”118

4.5.1 “可靠的估算工具”的使用方法118

4.5.2 “可靠的估算工具”的使用演练121

4.6 小结125

第二部分 分析需求

第5章 主体域分析128

5.1 本章概览128

5.2 什么是主体域130

5.3 主体域检查单131

5.3.1 “主体域检查单”的使用方法133

5.3.2 “主体域检查单”的使用演练139

5.4 主体域CRUD矩阵140

5.4.1 “主体域CRUD矩阵”的使用方法	142
5.4.2 “主体域CRUD矩阵”的使用演练	142
5.5 “In-the-Know模板”的使用	146
5.5.1 “In-the-Know模板”的使用方法	147
5.5.2 “In-the-Know模板”的使用演练	148
5.6 主体域家族树	148
5.6.1 “主体域家族树”的使用方法	151
5.6.2 “主体域家族树+”的使用方法	154
5.6.3 “主体域家族树”的使用演练	154
5.6.4 “主体域家族树+”的使用演练	158
5.7 主体域粒度矩阵	159
5.7.1 “主体域粒度矩阵”的使用方法	160
5.7.2 “主体域粒度矩阵”的使用演练	161
5.8 小结	166
第6章 主体域建模	167
6.1 本章概览	167
6.2 什么是主体域模型	168
6.3 如何解读规则	172
6.3.1 解读规则的实际演练	173
6.3.2 关于标注的忠告	174
6.3.3 关于基数的忠告	175
6.4 项目场景	175
6.5 什么是“业务清理板”模型	176
6.5.1 “业务清理板”模型的使用方法	177
6.5.2 “业务清理板”模型的使用演练	178
6.6 什么是“应用程序清理板”模型	182
6.6.1 “应用程序清理板”模型的使用 方法	184
6.6.2 “应用程序清理板”模型的使用 演练	184
6.7 什么是“早期现实性检查”模型	186
6.7.1 “早期现实性检查”模型的使用 方法	188
6.7.2 “早期现实性检查”模型的使用 演练	188
6.8 小结	191

第7章 逻辑数据分析	192
7.1 本章概览	192
7.2 什么是“数据元素家族树”	196
7.2.1 “数据元素家族树”的使用方法	199
7.2.2 “数据元素家族树”的使用演练	204
7.3 什么是“数据元素粒度矩阵”	215
7.3.1 “数据元素粒度矩阵”的使用方法	216
7.3.2 “数据元素粒度矩阵”的使用演练	218
7.4 什么是“数据质量记录模板”	219
7.4.1 “数据质量记录模板”的使用方法	225
7.4.2 “数据质量记录模板”的使用演练	228
7.5 什么是“数据质量确认模板”	235
7.5.1 “数据质量确认模板”的使用方法	236
7.5.2 “数据质量确认模板”的使用演练	237
7.6 小结	240

第三部分 建立需求模型及相关建议

第8章 规范化之旅和反向规范化生 存指南	242
8.1 本章概览	242
8.2 什么是规范化	245
8.3 什么是“规范化之旅”	246
8.4 什么是反向规范化	268
8.5 什么是“反向规范化生存指南”	274
8.5.1 “反向规范化生存指南”所包含 的问题	275
8.5.2 “反向规范化生存指南”的使用 方法	281
8.6 小结	288
第9章 抽象化安全指南与组件	290
9.1 本章概览	290
9.2 什么是抽象化	291
9.2.1 抽象化的好处	293
9.2.2 抽象化的代价	294
9.2.3 使用抽象化	296
9.3 什么是下级类型结构	300
9.4 什么是“抽象化安全指南”	302

9.4.1 共性	303	10.5 什么是“关系布局技巧”	361
9.4.2 目标	303	10.5.1 “关系布局技巧”的使用方法	362
9.4.3 工作量	305	10.5.2 “关系布局技巧”的使用演练	362
9.4.4 “抽象化安全指南”的使用方法	306	10.6 什么是“吸引注意力的技巧”	363
9.4.5 “抽象化安全指南”的使用演练	309	10.6.1 “吸引注意力的技巧”的使用方法	364
9.5 什么是“抽象组件”	332	10.6.2 “吸引注意力的技巧”的使用演练	365
9.5.1 使用实体抽象组件	332	10.7 小结	367
9.5.2 使用关系抽象组件	338	第11章 规划一个长盛不衰的数据建模生涯	368
9.5.3 使用数据元素抽象组件	340	11.1 本章概览	369
9.6 小结	344	11.2 对数据建模者的十大忠告清单	369
第10章 数据模型美化技巧	345	11.2.1 记住: 灵活性、准确性和背景	370
10.1 本章概览	345	11.2.2 建模只是你的工作的一小部分	371
10.2 什么是“逻辑数据元素排序技巧”	346	11.2.3 尝试其他角色	371
10.2.1 “逻辑数据元素排序技巧”的使用 方法	348	11.2.4 了解95/5规则	372
10.2.2 “逻辑数据元素排序技巧”的使用 演练	350	11.2.5 数据建模从不令人厌烦	373
10.3 什么是“物理数据元素排序技巧”	354	11.2.6 站在技术前沿	373
10.3.1 “物理数据元素排序技巧”的 使用方法	355	11.2.7 尽量不要在模型上牵扯上感情 因素	374
10.3.2 “物理数据元素排序技巧”的 使用演练	356	11.2.8 让你的创造力展开翅膀	374
10.4 什么是“实体布局技巧”	356	11.2.9 单纯的理论太昂贵了	375
10.4.1 “实体布局技巧”的使用方法	358	11.2.10 成为一个了不起的会讲故事的人	375
10.4.2 “实体布局技巧”的使用演练	358	11.3 小结	375
		推荐读物	376

第一部分 建立基础

你是否曾经遇到这样的处境：由于缺乏与项目经理的有效沟通，致使作为数据建模者的你做出不切实际的预期？给完成建模工作强行规定一个不切实际的时限就是这样一个例子，另外一个这方面的例子是：项目团队感到，许多数据建模活动是不必要的，简直是浪费时间。项目经理、数据建模者以及项目团队的其余成员需要对数据建模活动有现实而且相互可接受的预期。预期必须是现实的，也就是说，哪些建模任务是要求进行的，应该花多长时间来完成这些任务。预期必须是相互可接受的，也就是说，项目团队的所有成员，包括数据建模者，都要了解建模者的任务和时间约定，并就此达成共识。

现实而相互可接受的预期是不会自动出现的，仅仅通过积极的思考也不会实现这样的预期。作为数据建模者这样一个角色，你具备教育和影响他人的能力和职责。这些沟通方面的前提形成了应用程序的分析和设计活动的基础。在设计工作开始之前准备好这个基础，将增进项目经理对整个建模活动的满意度和赏识度。因此，这个基础对一个成功的建模活动来说极为关键。

本书的第一部分提供几个工具，以帮助作为数据建模者的你利用数据建模的教育和沟通方法。第1章到第4章包含了多个工具，它们重点关注数据建模教育和沟通的各个方面。以下是每一章的目的：

清晰地解释数据建模的概念。第1章关注我已经发现的、清晰表达数据建模的各个方面的、最有效的方式。趣闻（anecdote）就是一些小故事，它们能够很好地解释某个概念或术语。类比（analogy）把数据建模概念与读者非常熟悉的某个东西做对比，以强调数据建模的某些特性。演示文稿的制作步骤（presentation step）就是指准备一个数据建模演示文稿时经历的、有一定顺序的步骤。

对需要的元数据的类型达成共识。第2章有助于你为项目和组织定义标准的元数据类型。“元数据宾果”（Meta Data Bingo）是一种游戏，在这个游戏中，项目团队成员完成宾果卡片，识别出哪些类型的元数据最为重要，需要记录下来。这些宾果卡片记录多种不同细节程度的元数据，包括项目级的、数据模型级的、主体域级的、实体级的和数据元素级的。

改进定义的质量和一致性。第3章重点关注一个名为“定义检查单”（Definition Checklist）的工具，而且给出了在主体域级、实体级和数据元素级使用“定义检查单”的一系列实例。

提出一套数据建模任务，并阐明这些任务适用于较大的开发生命周期的方式。第4章包括四个项目计划模板，这些模板包含了一套完整而有序的任务，这些任务是在一个现实的期限内成功完成数据建模工作产品（deliverable）所不可或缺的。

第1章 使用趣闻、类比和演示文稿 来阐明数据建模的概念

曾有一次在欧洲度假时，我们决定在意大利北部一个很小的城镇中留连几日。这个小镇风景如画，周围是白雪盖顶的群山，山中到处是远足的小径。这个小镇中只有四家饭馆，我们决定尝试的第一家饭馆位于一个具有数百年历史的建筑物中。一进门，我们就立刻被这家饭馆的氛围和特色打动了——有许多古董和绘画，它们混合着壁炉和食物散发出的暖融融的气味。落座之后，我们发现这家饭馆里没有菜单。似乎任何时令的或者现有的食物都会被端上餐桌。这一般不会有什问题，除非这个饭馆中没有人会说英语；但我认识的惟一个意大利词语就是“ciao”，意思是“再见”。

我们需要一种交流方式，来清楚地表达我们的就餐愿望。尝试着和服务员进行几分钟不成功的、令人灰心丧气的交流之后，一个想法在我们脑海出现了。我们开始学着发出动物的叫声。显然，动物的叫声在各个地方都是一样的。通过发出牛叫的声音，我们就能够询问当晚是否供应牛肉。我们还可以发出猪叫的声音，同时摇着头，表示我们不想要猪肉。我们成功地扮演了所有的动物，准确地吃上了我们期望的食物。我很惊异这个技巧如此有效。通过模拟动物叫声的方式，我们能够在彼此之间进行准确地沟通。我认为，这次就餐是一个很好的例子，它说明，不管人们的背景和关注点如何不同，总会存在让大家相互进行准确地沟通的方法。

对许多应用程序开发经验有限的人来说，数据建模这个领域有时可能很陌生，就像一门外语一样。因此，他们可能对数据建模概念和术语缺乏了解，因而会对建模工作产品的完成产生混乱而不切实际的预期。通过清楚地解释数据建模的各个不同的方面，我们能够避免这种混乱的发生。有时，为了成功地解释数据建模概念，需要有创意的方法，就像在饭馆中模仿一头牛一样。我们需要想出鲜活的、创造性的方法，以便传达我们的思想。有时，一个小故事或者一个趣闻就能够清晰地解释一个数据建模概念；有时，类比也可能成为一个有效的技巧；有时，一个简短而有针对性的演示文稿也足以达到目的。我发现，趣闻、类比和演示文稿，使用这三种方式之一，或者混合使用这三种方式，就能够与难以掌握数据建模概念和术语的那些人们进行有效地沟通，有时还能影响他们的思想。

1.1 本章概览

本章重点关注一些最有效的工具，人们利用这些工具清晰地沟通数据建模的概念。在项目中和我们一起工作的某些人，如管理人员和业务用户，可能难以掌握诸如数据模型或数据仓库这样的概念。想像一下，向这些人机械地背诵数据模型的定义会很有效吗？他们可能会目瞪口呆，尽管他们也可能偶尔点点头，希望自己明白一点，但更有可能的是，他们将心烦意乱，甚至比以前更加糊涂。我们所需要的是一个简单而引人入胜的方法，用来解释建模的概念。

有时，会有人问我，我是以何为生的。作为一个数据建模者，你会怎样作答？因为发问的人通常没有太多的技术背景，所以，当我自豪地说“我是一个数据建模者”时，通常看到的是满脸的茫然。我可能像下面这样回答，“我给一个系统设计数据需求的表达方式，随后，这个系

统会被转化为一个关系数据库，实际的数据就会被输入到数据库中，并从数据库获取，等等。”有多少人真正理解或欣赏这样的解释？或者，我可以想到一个简单而有趣的方式来解释我的工作，我说“我为应用程序制订设计蓝图”，或者“我确保应用程序包含企业需要的、正确而完整的信息”。

改进人们之间的沟通方法的工具是存在的。在本章中，我们将讨论三种数据建模沟通工具：

趣闻。解释一个概念或术语时，使用小故事非常有效。我们许多人都喜欢听故事。我们期待着知晓接下来将发生什么事情。你可以结合使用幽默的小段、真实的故事和恰当的情节，来清晰地解释一个复杂的论题。我们将仔细查看建立一个数据建模故事的步骤，并且亲身体验一个详细的实例，这个实例是我曾经多次使用的一个真实故事。

类比。类比是两个或多个概念之间的对比。类比能够把听众熟悉的某个事物与我们希望他们了解的一个数据建模概念相比较。我们将讨论一个工作表，以利于创建一个有效的类比，而后，我们还要亲身体验几个我最喜欢的关于数据建模的类比。我们将描述其中六个关于数据建模的类比。因为我曾常常使用这些类比，而且它们也很有效，所以，在本章中，我将在这方面花费相当多的篇幅。

演示文稿的制作步骤。你曾经常常看到这样的演示文稿吗？你看了这个演示文稿，但你对它没有兴趣，因为你无法理解，它与你有何相关之处。当我们利用正式的演示文稿来解释数据建模概念时，我们需要确保，我们的资料已经经过裁剪或修改（定制），这样，这个演示文稿才能够深深打动听众的心扉，才能使表达出的信息对他们有直接影响。我们将讨论准备一个演示文稿的步骤，在准备这样的演示文稿时，我们会根据听众的情况修改信息。结构良好的演示文稿能够利用听众关注和欣赏的内容，来清晰地传达适当详细程度的必要的概念。我们将讨论准备这些定制化演示文稿的步骤，并亲身体验一个详细的实例。

如果你访想与本书相配套的网站www.wiley.com/compbooks/hoberman，将会找到本章中各个工具的模版、更多的工具，以及更多的关于本书中的工具的信息，包括你可以下载的空白模板。

1.2 什么是数据建模趣闻

在一般的日常沟通中，我们可能会说出并听到许多故事，或者趣闻，这些故事涉及的论题范围很大。有些例子是周末发生在我们自己身边的事情，或者是发生在我们认识的人身边的故事，或者是与我们的工作项目有关的经历。这些趣闻有助于加强我们和周围人们的关系，增进我们的愉悦情绪，而且对我们有教育作用。我们能够把由语言表达出来的东西形象化。有时，当故事结束时，给我们留下的是以前未曾想到的信息或更多的认识。在解释数据建模概念时，趣闻是极其有效的，原因有如下几个：

它们建立起持久的形象。与一个事实、一个论断或一个解释相比，我们的头脑更容易接受和牢记一个故事。例如，在一次招聘面谈中，如果你只告诉招聘人员你工作很努力，这是不会有持久的影响力的。你还不如讲述一个场合或一个故事，描述你怎样不分昼夜、放弃休假地工作，并最终完成一项任务。招聘人员将记住你不分昼夜、放弃休假工作的形象，这样建立的印象比单纯听你说你是一个工作努力的人而建立的印象要持久得多。有许多教授大家如何改进自