

数据模型记分卡

DATA MODEL SCORECARD

[美] 斯蒂夫·霍伯曼 (Steve Hoberman) 著 包春霞 译



数据模型记分卡

DATA MODEL SCORECARD

[美] 斯蒂夫·霍伯曼 (Steve Hoberman) 著 包春霞 译



人民邮电出版社
北京

图书在版编目 (C I P) 数据

数据模型记分卡 / (美) 霍伯曼 (Steve Hoberman)
著 ; 包春霞译. — 北京 : 人民邮电出版社, 2018.3
ISBN 978-7-115-47544-2

I. ①数… II. ①霍… ②包… III. ①数据模型—研
究 IV. ①TP311.13

中国版本图书馆CIP数据核字(2017)第315456号

版 权 声 明

Simplified Chinese translation copyright ©2018 by Posts and Telecommunications Press ALL RIGHTS RESERVED

Data Model Scorecard, by Steve Hoberman, ISBN 9781634620826

Copyright © 2015 by Technics Publications, LLC

本书中文简体版由 **Technics Publications** 授权人民邮电出版社出版。未经出版者书面许可, 对本书的任何部分不得以任何方式或任何手段复制和传播。

版权所有, 侵权必究。

-
- ◆ 著 [美] 斯蒂夫·霍伯曼 (Steve Hoberman)
译 包春霞
责任编辑 胡俊英
责任印制 焦志炜
- ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
大厂聚鑫印刷有限责任公司印刷
- ◆ 开本: 720×960 1/16
印张: 11.5
字数: 152 千字 2018 年 3 月第 1 版
印数: 1—2 000 册 2018 年 3 月河北第 1 次印刷
- 著作权合同登记号 图字: 01-2016-3955 号
-

定价: 59.00 元

读者服务热线: (010)81055410 印装质量热线: (010)81055316

反盗版热线: (010)81055315



内容提要

数据建模是对现实世界各类数据进行抽象组织、界定数据库需管辖的范围、确定数据的组织形式等直至转化成现实数据库的过程。而数据模型是构建应用系统的核心，是尽可能精准地表示业务运转的概念性框架。

本书重在介绍数据建模领域的经典理论——数据模型记分卡。全书通过 16 章内容，分 3 个部分介绍了如何使用数据模型记分卡来改进数据模型。第 1 部分介绍了数据建模和验证；第 2 部分介绍了数据模型记分卡的分类，列举了 10 种数据模型记分卡；第 3 部分介绍了记分卡在模型评审中的使用。

作为一本经典大师级著作，本书非常适合对数据建模感兴趣的读者以及从事数据库等相关工作的专业人士参考阅读。通过阅读本书，读者将对记分卡这一经典理论有更加全面、深入的理解。



前言

数据建模的过程是界定、分析、发现数据需求，再用可视化的形式（“模型”）表示这种数据需求的过程。数据模型是用于精确表示信息领域沟通的一套符号和文字。任何景观的模型都会包含某些内容（例如地图就是地理景观的模型），同时为了方便理解又排除某些内容。

“发现”是确定业务过程或应用中业务需要什么信息，例如了解到客户和账号是个重要的概念。“分析”是明确需求的过程，例如对客户和账户逐步有了清楚的定义，理解了客户与他们的账户之间的关系。“界定范围”涉及与业务合作来决定什么对于特定的业务阶段是最重要的。例如，第一阶段是否同时需要“保存”和“检查账号”，还是只要“检查账号”就行了。“表示”是指要用清晰明确的语言展现出信息景观看上去是什么样的，例如可以用以下数据模型表示：



- 一个客户可以有一个或多个账号。
- 一个账号必须由一个或多个客户拥有。

一旦我们将这些需求写成了数据模型文档，就可以跟应用开发所涉及的业务和信息技术（Information Technology, IT）人员进行沟通了，如业务用户、业务分析师、数据建模人员、数据架构师、数据库管理员、开发者、测试人员以及管理人员。

数据模型是用于从业务到 IT，IT 内部从系统分析员、建模人员、架构

师到数据库设计人员和开发人员之间沟通的主要媒介。无论要用的数据库技术是关系数据库管理系统（Relational Database Management System, RDBMS）（如 ORACLE、Teradata），还是像 MongoDB 或 Hadoop 这样的非关系型数据库（Not Only SQL, NoSQL），都需要有某种方式能用来沟通数据需求。因此，我们需要数据模型。

数据模型应该是高质量的，要能支持目前的需求同时又要能满足未来的需要。数据模型记分卡是一个可以用来改进数据模型质量的工具。

许多我正在为他们提供咨询服务的客户都决定将数据记分卡应用到客户的数据模型中，他们也推荐用数据模型记分卡来改进设计。本书将为读者展示如何应用数据模型记分卡。本书正是为那些构造、使用或评审数据模型的人员写的，全书分为 3 个部分。

第 1 部分，数据建模和验证。

- 第 1 章简单介绍数据建模；
- 第 2 章讲解为何正确的数据模型如此重要；
- 第 3 章介绍数据模型记分卡。

第 2 部分，数据模型记分卡分类，我们将对 10 种数据模型记分卡分别进行探索。这部分有 10 章，每章分别对一种记分卡进行介绍。

- 第 4 章 分类 1：正确性
- 第 5 章 分类 2：完整性
- 第 6 章 分类 3：模式
- 第 7 章 分类 4：结构
- 第 8 章 分类 5：抽象
- 第 9 章 分类 6：标准
- 第 10 章 分类 7：可读性
- 第 11 章 分类 8：定义
- 第 12 章 分类 9：一致性

- 第 13 章 分类 10: 数据

每一章都以对该分类的检查总结作为结束。

第 3 部分，用记分卡验证数据模型，这部分将为模型评审（详见第 14 章）做准备，提出了一些关于模型评审的建议（详见第 15 章），接着对一个实际项目做数据模型评审（详见第 16 章）。

本书中出现的所有数据模型都是用 Embarcadero 公司的数据建模工具 ER/Studio 创建的。



目录

第1部分 数据建模和验证

第1章 数据模型基础	2
1.1 实体	2
1.2 属性	5
1.3 域	5
1.4 关系	6
1.5 键	9
1.6 代理键	13
1.7 外键	15
1.8 次键	16
1.9 子类型	16
第2章 数据模型质量的重要性	19
2.1 精确性	19
2.2 杠杆	21
2.3 数据质量	21
第3章 数据模型记分卡概述	23
3.1 记分卡的特征	23
3.2 记分卡模板	24
3.3 DMM 上下文	29

3.3.1 DMM 和数据建模	32
3.3.2 功能实践陈述	34
3.4 DMBOK 上下文	37

第 2 部分 数据模型记分卡分类

第 4 章 分类 1: 正确性	42
4.1 模型是否正确地捕获了需求	42
4.2 分类的期望	43
4.3 对正确性检查的小结	50
第 5 章 分类 2: 完整性	51
5.1 模型的完整性如何	51
5.2 类别期望	51
5.2.1 确认获得所有的业务元数据	52
5.2.2 确认获得所有的技术元数据	52
5.2.3 确认每个需求都被完全捕获	53
5.2.4 确认没有提供免费的服务	53
5.2.5 确认澄清了所有模糊的需求	54
5.3 对完整性检查的小结	55
第 6 章 分类 3: 模式	56
6.1 模型与其模式匹配得如何	56
6.2 类型期望	61
6.2.1 概念模型遵从	61
6.2.2 逻辑模型遵从	67
6.2.3 物理模型遵从	83
6.3 模式检查的小结	93

第 7 章 分类 4: 结构	97
7.1 分类期望	97
7.1.1 模型是一致的	97
7.1.2 模型有完整性	99
7.1.3 模型遵从核心原则	103
7.2 结构检查的小结	106
第 8 章 分类 5: 抽象	107
8.1 模型是否很好地利用了通用的结构	107
8.2 类型期望	108
8.2.1 模型是可扩展的	108
8.2.2 模型是可用的	110
8.3 抽象检查的小结	111
第 9 章 分类 6: 标准	112
9.1 模型是否很好地遵从命名规范	112
9.2 分类期望	112
9.2.1 模型要结构良好	113
9.2.2 模型使用正确的术语	117
9.2.3 模型有一致的样式	119
9.3 标准检查的小结	119
第 10 章 分类 7: 可读性	121
10.1 模型是否具有可读性	121
10.2 分类期望	121
10.2.1 模型是可读的	121
10.2.2 实体布局可接受	124
10.2.3 属性顺序可接受	127
10.2.4 关系布局可接受	128

10.3 可读性检查的小结	128
第 11 章 分类 8: 定义	130
11.1 定义是否做得足够好	130
11.2 分类期望	131
11.2.1 定义是清晰的	131
11.2.2 定义是完整的	132
11.2.3 定义是正确的	133
11.3 定义检查的小结	134
第 12 章 类别 9: 一致性	135
12.1 模型与企业的一致性如何	135
12.2 分类期望	136
12.3 一致性检查的小结	137
第 13 章 分类 10: 数据	138
13.1 元数据与数据匹配得如何	138
13.2 分类期望	138
13.3 数据检查的小结	141

第 3 部分 用记分卡验证数据模型

第 14 章 准备模型评审	144
14.1 必需的文档	144
14.2 附加的信用文档	145
14.3 评审结构	146
14.3.1 将模型分块评审	146
14.3.2 决定谁参加评审	147
14.3.3 座位模式	147

第 15 章 模型评审	149
15.1 一米深胜过一里宽	149
15.2 设置阶段	150
15.3 建立一个支持小组	150
15.4 平均分数	151
15.5 从 CDM 开始	151
15.6 知道何时停止	151
15.7 保持有趣	151
第 16 章 数据模型记分卡案例研究：消费者互动	152
16.1 正确性	155
16.2 完整性	156
16.3 模式	157
16.4 结构	159
16.5 抽象	160
16.6 标准	160
16.7 可读性	161
16.8 定义	162
16.9 一致性	165
16.10 数据	166
16.11 完成的记分卡	168



第 1 部分

数据建模和验证



在这一部分，第 1 章是一个简短的数据建模介绍，理解为什么要用数据模型，第 2 章介绍如何获得一个正确的数据模型，第 3 章介绍数据模型记分卡。



第 1 章

数据模型基础

本章将简要介绍数据模型的组成，并教你如何看懂数据模型。更多的内容请参考我的另外一本书《数据建模经典教程》(Data Modeling made Simple)。

1.1 实体

实体表示与业务有关的重要且有价值的事务的信息集合。每个实体由一个名词或名词词组来表示，一般适用于以下 6 种问题之一：谁、什么、何时、哪里、为什么、如何。表 1-1 是这些实体类的定义并举例说明。

表 1-1 实体类的定义

分类	定义	举例
谁 (Who)	能为企业带来好处的个人或组织。“谁对业务很重要？”常常与角色有关，例如客户或供应商	员工、病人、演员、嫌疑人、客户、供应商、学生、旅客、参赛者、作者
什么 (What)	对企业有利的产品或服务。常常指能使组织保持业务运转的产出物。“什么是对业务至关重要的？”	产品、服务、原材料、货物清单、课程、歌曲、照片、图书
何时 (When)	企业所关心的日历或时间周期。“何时业务在运作？”	时间、日期、月份、季度、年、学期、会计期间、分钟
哪里 (Where)	企业关心的位置。位置可以指实际的地理位置，也可以指虚拟的位置。“业务在哪里开展？”	邮件地址、分布地点、网站的 URL 及 IP 地址
为什么 (Why)	企业所关心的事件或事务。事件会让业务具有不确定性。“业务为什么可以运转？”	订单、盈利、投诉、取钱、存钱、褒扬、问询、交易、索赔

续表

分类	定义	举例
如何 (How)	将企业关心的事件记录下来。可以用文档记录事件,如采购订单用于记录一个订单事件。这就是“业务是如何跟踪事件的?”	发票、合同、协议、购买订单、收据、发票、装箱清单、交易确认单

实体的实例是指特定实体的发生或实体的值。例如,表单就是一个实体,它的表头的每个字段表示每个实体要记录的信息。每个有实际值的表单行表示一个实体实例。“客户”实体可能会有多个有不同名字的客户实例,如 Bob、Joe、Jane 等。“账号”实体有 Bob 的支票账户实例、Bob 的存款账户实例、Joe 的佣金账户实例等。

实体可以从概念层、逻辑层、物理层进行描述。概念层是对一个业务过程或应用系统定义其范围和重要术语。逻辑层是对一个业务过程或应用系统的业务解决方案进行详细描述,物理层则是对一个应用系统的技术解决方案进行详细描述。

一个与概念层相关的实体一定是对业务基本且关键的。至于什么是“基本且关键的”主要因范围不同而不同。在通用层面,某些概念是大多数公司都共有的,例如,客户、产品和员工。稍微收窄这个范围,特定的行业可能会有某个特定的概念。例如,广告役,这在广告行业是有一个有效的概念,但在其他行业中却不适用。在出版行业,作者 (Author)、书 (Title) 和订单 (Order) 是概念实体,如图 1-1 矩形框中的名字。



图 1-1 用矩形框中的名字表示概念实体

逻辑层的实体比概念层更加详细地表示了业务。经常会用一个概念实体表示多个逻辑实体。逻辑实体包含一些特性,我们叫作“属性”,下一节会讨论。前面的概念实体可以由 3 个逻辑实体表示,如图 1-2 所示。

Author	Title	Order
Author Tax Identifier	Title ISBN	Order Number
Author Birth Date	Title Name	Order Placement Date
Author Last Name	Subtitle Name	Order Scheduled Delivery Date
Author First Name	Retail Price Amount	

图 1-2 逻辑实体

在物理层，实体与技术特定的对象有关，例如关系数据库管理系统（Relational Database Management System, RDBMS）中的数据库表或者非关系型数据库（Not Only SQL, NoSQL）MongoDB 中的集合。物理层与逻辑层相似，但可以包括弥补技术缺陷所需要的折中方案，一般是与性能或存储有关的。

以下是前述逻辑实体的物理实体，如图 1-3 所示。物理实体包含数据库特定的信息，例如，属性的格式和长度（作者的姓氏，即 Author Last Name 是 50 个字符长度），该属性是否必须有值（作者的税号标识，即 Author Tax Identifier 不能为空因而必须有值，作者的出生日期，即 Author Birth Date 是可空的，所以不要求一定有值）。

Author	Title	Order
Author Tax Identifier (PK) CHAR(9) NOT NULL	Title ISBN (PK) CHAR(13) NOT NULL	Order Number (PK) CHAR(5) NOT NULL
Author Birth Date DATETIME NULL	Title Name CHAR(100) NULL	Order Placement Date DATE NULL
Author Last Name CHAR(50) NULL	Subtitle Name CHAR(100) NULL	Order Scheduled Delivery Date DATE NULL
Author First Name CHAR(30) NULL	Retail Price Amount DECIMAL(5,2) NULL	

图 1-3 物理实体

在关系数据库管理系统（Relational Database Management System, RDBMS）中，这些物理实体是数据库表或视图。在非关系型数据库（Not Only SQL, NoSQL）中，这些物理实体根据所用的技术不同而不同。例如，在 MongoDB 这样的文档数据库中，这些实体是集合。常用的术语“结构”是指数据库组件，无论这种数据库是 RDBMS 还是 NoSQL 类型。

1.2 属性

属性是用来识别、描述或度量实体实例的单个信息单元。货运单号（Claim Number）属性标识每个货运单。学生姓氏（Student Last Name）属性描述每个学生。订单金额（Gross Sales Amount）属性度量每笔交易的金额。

像实体一样，属性也可以在概念、逻辑和物理层描述。概念层的属性必须是对业务基本且关键的。通常我们并不从概念层描述属性，当然根据业务需求的不同，也可以作为概念层的属性。我在一家通讯公司工作时，电话号码是对业务非常重要的属性，所以多个概念模型中都有电话号码属性。

逻辑模型的属性表示业务特性。每个属性显示出它对业务解决方案的贡献，并且独立于任何软硬件技术。例如，作者的姓氏（Author Last Name）就是一个属性，因为它有重要的业务意义，无论这些记录是保存在纸质文件里还是能快速检索的数据库里。物理模型的属性代表数据库的一列。作者的姓氏（Author Last Name）属性可能在关系数据库的 AUTH 表里用 AUTH_LAST_NM 列表示，或者在 MongoDB 集合 LibraryCardCatalog 中用 AuthorLastName 表示。

1.3 域

一个属性所有可能赋值的全部集合叫作域。域包括一套可应用于不止一个属性的验证标准。例如，日期（Date）域含有可给以下这些属性赋值的所有可能的有效日期。

- 员工雇用日期
- 下单日期
- 交货日期
- 课程开始日期

一个属性一定不能包含其赋值域以外的值。域值由特定的实际列表值