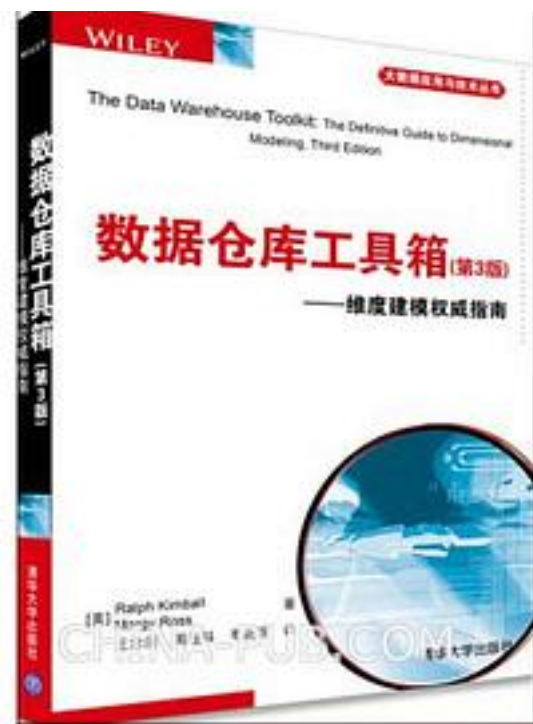


数据仓库工具箱（第3版）



[数据仓库工具箱（第3版） 下载链接1](#)

著者:Ralph Kimball

出版者:清华大学出版社

出版时间:2015-1-10

装帧:平装

isbn:9787302385530

随着The Data Warehouse Toolkit(1996)第1版的出版发行，Ralph Kimball为整个行业引入了维度建模技术。从此，维度建模成为一种被广泛接受的表达数据仓库和商业智能(DW/BI)系统中数据的方法。该经典书籍被认为是维度建模技术、模式和最佳实践的权威资源。

这本《数据仓库工具箱(第3版)——维度建模权威指南》汇集了到目前为止最全面的维度建模技术。本书采用新的思路和最佳实践对上一版本进行了全面修订，给出了设计维度模型的全面指南，既适合数据仓库新手，也适合经验丰富的专业人员。

《数据仓库工具箱(第3版)——维度建模权威指南》涉及的所有技术都基于作者实际从事DW/BI的设计经验，通过实际案例加以描述。

主要内容 ◆ 实用设计技术——有关维度和事实表的基本和高级技术

◆ 14个案例研究，涉及零售业、电子商务、客户关系管理、采购、库存、订单管理、会计、人力资源、金融服务、医疗卫生、保险、教育、电信和运输等

◆ 为12个案例研究提供了数据仓库总线矩阵示例

◆ 需要避免的维度建模陷阱和错误

◆ 增强的缓慢变化维度(SCD)技术类型0~类型7

◆ 用于处理参差不齐的可变深度层次和多值属性的桥接表

◆ 大数据分析的最佳实践

◆ 与业务参与方合作、交互设计会议的指南

◆ 有关Kimball DW/BI项目生命周期方法论的概论

◆ 对ETL系统和设计思考的总结

◆ 构建维度和事实表的34个ETL子系统和技術

作者介绍:

Ralph

Kimball是Kimball集团的创建者。从20世纪80年代中期以来，他一直是数据仓库和商业智能行业维度建模方法的思想开拓者。大量IT专业人士接受过其教育。自1996年以来，由他及其同事们所撰写的工具箱系列书籍一直是最受读者青睐的书籍。Ralph

Kimball曾就职于Metaphor并建立了Red Brick系统，他在施乐Palo Alto研究中心(PARC)工作期间，与他人一起共同发明了星型工作站，这是首个利用视窗、图标和鼠标的商业产品。Ralph

Kimball毕业于斯坦福大学电子工程系并获得博士学位。

Margy

Ross是Kimball集团总裁。自1982年以来，她主要关注数据仓库和商业智能，强调业务需求和维度建模的重要性。与Ralph Kimball一样，Margy

Ross也为许多学生讲授过维度设计最佳实践，她与Ralph

Kimball合作，共同撰写了5本工具箱序列书籍。Margy

Ross曾工作于Metaphor并与他人共同创立了DecisionWorks咨询公司。她毕业于美国西北大学工业工程系并获得硕士学位。

目录: 第1章 数据仓库、商业智能及维度建模初步 1

1.1 数据获取与数据分析的区别 1

1.2 数据仓库与商业智能的目标 2

1.3 维度建模简介 5

1.3.1 星型模式与OLAP多维数据库 6

1.3.2 用于度量的事实表 7

1.3.3 用于描述环境的维度表	9
1.3.4 星型模式中维度与事实的连接	11
1.4 Kimball的DW/BI架构	14
1.4.1 操作型源系统	14
1.4.2 获取-转换-加载(ETL)系统	14
1.4.3 用于支持商业智能决策的展现区	16
1.4.4 商业智能应用	17
1.4.5 以餐厅为例描述Kimball架构	17
1.5 其他DW/BI架构	19
1.5.1 独立数据集市架构	19
1.5.2 辐射状企业信息工厂Inmon架构	20
1.5.3 混合辐射状架构与Kimball架构	22
1.6 维度建模神话	22
1.6.1 神话1: 维度模型仅包含汇总数据	23
1.6.2 神话2: 维度模型是部门级而不是企业级的	23
1.6.3 神话3: 维度模型是不可扩展的	23
1.6.4 神话4: 维度模型仅用于预测	23
1.6.5 神话5: 维度模型不能被集成	24
1.7 考虑使用维度模型的更多理由	24
1.8 本章小结	25
第2章 Kimball维度建模技术概述	27
2.1 基本概念	27
2.1.1 收集业务需求与数据实现	27
2.1.2 协作维度建模研讨	27
2.1.3 4步骤维度设计过程	28
2.1.4 业务过程	28
2.1.5 粒度	28
2.1.6 描述环境的维度	28
2.1.7 用于度量的事实	29
2.1.8 星型模式与OLAP多维数据库	29
2.1.9 方便地扩展到维度模型	29
2.2 事实表技术基础	29
2.2.1 事实表结构	29
2.2.2 可加、半可加、不可加事实	29
2.2.3 事实表中的空值	30
2.2.4 一致性事实	30
2.2.5 事务事实表	30
2.2.6 周期快照事实表	30
2.2.7 累积快照事实表	30
2.2.8 无事实的事实表	31
2.2.9 聚集事实表或OLAP多维数据库	31
2.2.10 合并事实表	31
2.3 维度表技术基础	31
2.3.1 维度表结构	31
2.3.2 维度代理键	32
2.3.3 自然键、持久键和超自然键	32
2.3.4 下钻	32
2.3.5 退化维度	32
2.3.6 非规范化扁平维度	32
2.3.7 多层次维度	32
2.3.8 文档属性的标识与指示器	33
2.3.9 维度表中的空值属性	33
2.3.10 日历日期维度	33
2.3.11 扮演角色的维度	33

- 2.3.12 杂项维度 33
- 2.3.13 雪花维度 33
- 2.3.14 支架维度 34
- 2.4 使用一致性维度集成 34
 - 2.4.1 一致性维度 34
 - 2.4.2 缩减维度 34
 - 2.4.3 跨表钻取 34
 - 2.4.4 价值链 34
 - 2.4.5 企业数据仓库总线架构 35
 - 2.4.6 企业数据仓库总线矩阵 35
 - 2.4.7 总线矩阵实现细节 35
 - 2.4.8 机会/利益相关方矩阵 35
- 2.5 处理缓慢变化维度属性 35
 - 2.5.1 类型0：原样保留 35
 - 2.5.2 类型1：重写 35
 - 2.5.3 类型2：增加新行 36
 - 2.5.4 类型3：增加新属性 36
 - 2.5.5 类型4：增加微型维度 36
 - 2.5.6 类型5：增加微型维度及类型1支架 36
 - 2.5.7 类型6：增加类型1属性到类型2维度 36
 - 2.5.8 类型7：双类型1和类型2维度 36
- 2.6 处理维度层次关系 37
 - 2.6.1 固定深度位置的层次 37
 - 2.6.2 轻微参差不齐/可变深度层次 37
 - 2.6.3 具有层次桥接表的参差不齐/可变深度层次 37
 - 2.6.4 具有路径字符属性的可变深度层次 37
- 2.7 高级事实表技术 37
 - 2.7.1 事实表代理键 37
 - 2.7.2 蜈蚣事实表 38
 - 2.7.3 属性或事实的数字值 38
 - 2.7.4 日志/持续时间事实 38
 - 2.7.5 头/行事实表 38
 - 2.7.6 分配的事实 38
 - 2.7.7 利用分配建立利润与损失事实表 38
 - 2.7.8 多种货币事实 39
 - 2.7.9 多种度量事实单位 39
 - 2.7.10 年-日事实 39
 - 2.7.11 多遍SQL以避免事实表间的连接 39
 - 2.7.12 针对事实表的时间跟踪 39
 - 2.7.13 迟到的事实 40
- 2.8 高级维度技术 40
 - 2.8.1 维度表连接 40
 - 2.8.2 多值维度与桥接表 40
 - 2.8.3 随时间变化的多值桥接表 40
 - 2.8.4 标签的时间序列行为 40
 - 2.8.5 行为研究分组 40
 - 2.8.6 聚集事实作为维度属性 41
 - 2.8.7 动态值范围 41
 - 2.8.8 文本注释维度 41
 - 2.8.9 多时区 41
 - 2.8.10 度量类型维度 41
 - 2.8.11 步骤维度 41
 - 2.8.12 热交换维度 42
 - 2.8.13 抽象通用维度 42

2.8.14 审计维度	42
2.8.15 最后产生的维度	42
2.9 特殊目的模式	42
2.9.1 异构产品的超类与子类模式	43
2.9.2 实时事实表	43
2.9.3 错误事件模式	43
第3章 零售业务	45
3.1 维度模型设计的4步过程	46
3.1.1 第1步：选择业务过程	46
3.1.2 第2步：声明粒度	46
3.1.3 第3步：确定维度	47
3.1.4 第4步：确定事实	47
3.2 零售业务案例研究	47
3.2.1 第1步：选择业务过程	49
3.2.2 第2步：声明粒度	49
3.2.3 第3步：确定维度	50
3.2.4 第4步：确定事实	50
3.3 维度表设计细节	53
3.3.1 日期维度	53
3.3.2 产品维度	56
3.3.3 商店维度	59
3.3.4 促销维度	60
3.3.5 其他零售业维度	62
3.3.6 事务号码的退化维度	63
3.4 实际的销售模式	63
3.5 零售模式的扩展能力	64
3.6 无事实的事实表	65
3.7 维度与事实表键	66
3.7.1 维度表代理键	66
3.7.2 维度中自然和持久的超自然键	68
3.7.3 退化维度的代理键	68
3.7.4 日期维度的智能键	68
3.7.5 事实表的代理键	69
3.8 抵制规范化的冲动	70
3.8.1 具有规范化维度的雪花模式	70
3.8.2 支架表	72
3.8.3 包含大量维度的蜈蚣事实表	72
3.9 本章小结	74
第4章 库存	75
4.1 价值链简介	75
4.2 库存模型	76
4.2.1 库存周期快照	76
4.2.2 库存事务	79
4.2.3 库存累积快照	80
4.3 事实表类型	81
4.3.1 事务事实表	81
4.3.2 周期快照事实表	82
4.3.3 累积快照事实表	82
4.3.4 辅助事实表类型	83
4.4 价值链集成	83
4.5 企业数据仓库总线架构	84
4.5.1 理解总线架构	84
4.5.2 企业数据仓库总线矩阵	85
4.6 一致性维度	89

4.6.1 多事实表钻取	89
4.6.2 相同的一致性维度	89
4.6.3 包含属性子集的缩减上卷一致性维度	90
4.6.4 包含行子集的缩减一致性维度	91
4.6.5 总线矩阵的缩减一致性维度	91
4.6.6 有限一致性	92
4.6.7 数据治理与管理的重要性	92
4.6.8 一致性维度与敏捷开发	94
4.7 一致性事实	94
4.8 本章小结	95
第5章 采购	97
5.1 采购案例研究	97
5.2 采购事务与总线矩阵	98
5.2.1 单一事务事实表与多事务事实表	98
5.2.2 辅助采购快照	101
5.3 缓慢变化维度(SCD)基础	101
5.3.1 类型0: 保留原始值	102
5.3.2 类型1: 重写	102
5.3.3 类型2: 增加新行	104
5.3.4 类型3: 增加新属性	106
5.3.5 类型4: 增加微型维度	108
5.4 混合缓慢变化维度技术	110
5.4.1 类型5: 微型维度与类型1支架表	110
5.4.2 类型6: 将类型1属性增加到类型2维度	111
5.4.3 类型7: 双重类型1与类型2维度	112
5.5 缓慢变化维度总结	113
5.6 本章小结	114
第6章 订单管理	115
6.1 订单管理总线矩阵	116
6.2 订单事务	116
6.2.1 事实表规范化	117
6.2.2 维度角色扮演	117
6.2.3 重新审视产品维度	119
6.2.4 客户维度	120
6.2.5 交易维度	122
6.2.6 针对订单号的退化维度	123
6.2.7 杂项维度	124
6.2.8 应该避免的表头/明细模式	125
6.2.9 多币种	126
6.2.10 不同粒度的事务事实	128
6.2.11 另外一种需要避免的表头/明细模式	129
6.3 发票事务	130
6.3.1 作为事实、维度或两者兼顾的服务级性能	131
6.3.2 利润与损益事实	131
6.3.3 审计维度	133
6.4 用于订单整个流水线的累积快照	134
6.4.1 延迟计算	136
6.4.2 多种度量单位	137
6.4.3 超越后视镜	138
6.5 本章小结	138
第7章 会计	139
7.1 会计案例研究与总线矩阵	139
7.2 总账数据	141
7.2.1 总账周期快照	141

7.2.2 会计科目表	141
7.2.3 结账	141
7.2.4 年度-日期事实	143
7.2.5 再次讨论多币种问题	143
7.2.6 总账日记账事务	143
7.2.7 多种财务会计日历	144
7.2.8 多级别层次下钻	145
7.2.9 财务报表	145
7.3 预算编制过程	146
7.4 维度属性层次	148
7.4.1 固定深度的位置层次	148
7.4.2 具有轻微不整齐的可变深度层次	149
7.4.3 不整齐可变深度层次	149
7.4.4 不规则层次中的共享所有权	152
7.4.5 随时间变化的不规则层次	153
7.4.6 修改不规则层次	153
7.4.7 其他不规则层次的建模方法	154
7.4.8 应用于不规则层次的桥接表方法的优点	156
7.5 合并事实表	156
7.6 OLAP角色及分析方案包	157
7.7 本章小结	158
第8章 客户关系管理	159
8.1 客户关系管理概述	160
8.2 客户维度属性	162
8.2.1 名字与地址的语法分析	162
8.2.2 国际姓名和地址的考虑	164
8.2.3 客户为中心的日期	165
8.2.4 作为维度属性的聚集事实	166
8.2.5 分段属性与记分	166
8.2.6 包含类型2维度变化的计算	169
8.2.7 低粒度属性集合的支架表	169
8.2.8 客户层次的考虑	170
8.3 应用于多值维度的桥接表	171
8.3.1 稀疏属性的桥接表	172
8.3.2 应用于客户多种联系方式的桥接表	173
8.4 复杂的客户行为	173
8.4.1 客户队列的行为研究分组	173
8.4.2 连续行为的步骤维度	175
8.4.3 时间范围事实表	176
8.4.4 使用满意度指标标记事实表	177
8.4.5 使用异常情景指标标记事实表	178
8.5 客户数据集成方法	178
8.5.1 建立单一客户维度的主数据管理	179
8.5.2 多客户维度的局部一致性	180
8.5.3 避免对应事实表的连接	180
8.6 低延迟的实现检查	181
8.7 本章小结	182
第9章 人力资源管理	183
9.1 雇员档案跟踪	183
9.1.1 精确的有效和失效时间范围	184
9.1.2 维度变化原因跟踪	185
9.1.3 作为类型2属性或事实事件的档案变化	185
9.2 雇员总数周期快照	186
9.3 人力资源过程的总线矩阵	187

9.4 分析解决方案软件包与数据模型	188
9.5 递归式雇员层次	189
9.5.1 针对嵌入式经理主键变化的跟踪	190
9.5.2 上钻或下钻管理层次	190
9.6 多值技能关键字属性	191
9.6.1 技能关键字桥接表	191
9.6.2 技能关键字文本字符串	192
9.7 调查问卷数据	193
9.8 本章小结	194
第10章 金融服务	195
10.1 银行案例研究与总线矩阵	195
10.2 分类维度以避免出现维度太少的情况	196
10.2.1 家庭维度	199
10.2.2 多值维度与权重因子	199
10.2.3 再谈微型维度	200
10.2.4 在桥接表中增加微型维度	202
10.2.5 动态值范围事实	202
10.3 异构产品的超类和子类模式	203
10.4 热可交换维度	205
10.5 本章小结	205
第11章 电信	207
11.1 电信业案例研究与总线矩阵	207
11.2 设计评审的一般性考虑	209
11.2.1 业务需求与实际可用资源的权衡	209
11.2.2 关注业务过程	209
11.2.3 粒度	210
11.2.4 统一的事实表粒度	210
11.2.5 维度的粒度和层次	210
11.2.6 日期维度	211
11.2.7 退化维度	211
11.2.8 代理键	212
11.2.9 维度解码与描述符	212
11.2.10 一致的承诺	212
11.3 设计评审指导	212
11.4 草案设计训练的讨论	214
11.5 重新建模已存在的数据结构	215
11.6 地理位置维度	216
11.7 本章小结	216
第12章 交通运输	217
12.1 航空案例研究与总线矩阵	217
12.1.1 多种事实表粒度	218
12.1.2 连接区段形成旅程	220
12.1.3 相关事实表	221
12.2 扩展至其他行业	221
12.2.1 货物托运人	221
12.2.2 旅行服务	222
12.3 相关维度合并	222
12.3.1 服务类别	223
12.3.2 始发地与目的地	224
12.4 更多有关日期和时间的考虑	225
12.4.1 用作支架表的特定国家日历	225
12.4.2 多时区的日期和时间	226
12.5 本地化概要	226
12.6 本章小结	227

第13章 教育	229
13.1 大学案例研究与总线矩阵	229
13.2 累积快照事实表	231
13.2.1 申请流水线	231
13.2.2 科研资助项目流水线	232
13.3 无事实的事实表	232
13.3.1 招生事件	233
13.3.2 课程注册	233
13.3.3 设施使用	235
13.3.4 学生考勤	236
13.4 更多关于教育分析的情况	237
13.5 本章小结	237
第14章 医疗卫生	239
14.1 医疗卫生案例研究与总线矩阵	239
14.2 报销单据与支付	241
14.2.1 日期维度角色扮演	243
14.2.2 多值诊断	243
14.2.3 收费的超类与子类	245
14.3 电子医疗记录	246
14.3.1 度量稀疏事实的类型维度	246
14.3.2 自由文本注释	247
14.3.3 图像	247
14.4 设施/设备的库存利用	247
14.5 处理可追溯的变化	248
14.6 本章小结	248
第15章 电子商务	249
15.1 点击流源数据	249
15.2 点击流维度模型	252
15.2.1 网页维度	252
15.2.2 事件维度	253
15.2.3 会话维度	254
15.2.4 推荐维度	254
15.2.5 点击流会话事实表	255
15.2.6 点击流网页事件事实表	256
15.2.7 步骤维度	258
15.2.8 聚集点击流事实表	258
15.2.9 Google Analytics(GA)	259
15.3 将点击流集成到Web零售商总线矩阵中	259
15.4 包含Web的跨渠道赢利能力	261
15.5 本章小结	263
第16章 保险业务	265
16.1 保险案例研究	266
16.1.1 保险业价值链	266
16.1.2 总线矩阵草案	267
16.2 保单事务	268
16.2.1 维度角色扮演	268
16.2.2 缓慢变化维度	268
16.2.3 针对大型和快速变化维度的微型维度	269
16.2.4 多值维度属性	269
16.2.5 作为事实或维度的数值属性	270
16.2.6 退化维度	270
16.2.7 低粒度维度表	270
16.2.8 审计维度	270
16.2.9 保单事务事实表	270

16.2.10 异构的超类和子类产品	271
16.2.11 辅助保险累积快照	272
16.3 保费周期快照	272
16.3.1 一致性维度	272
16.3.2 一致性事实	273
16.3.3 预付事实	273
16.3.4 再谈异构超类与子类	273
16.3.5 再谈多值维度	274
16.4 更多保险案例研究背景	274
16.4.1 更新保险行业总线矩阵	275
16.4.2 总线矩阵实现细节	275
16.5 索赔事务	277
16.6 索赔累积快照	278
16.6.1 复杂工作流的累积快照	279
16.6.2 时间范围累积快照	279
16.6.3 周期而不是累积快照	280
16.7 保单/索赔合并的周期快照	280
16.8 无事实的意外事件	280
16.9 需要避免的常见维度建模错误	281
16.9.1 错误10：在事实表中放入文本属性	281
16.9.2 错误9：限制使用冗长的描述符以节省空间	281
16.9.3 错误8：将层次划分为多个维度	282
16.9.4 错误7：忽略对维度变化进行跟踪的需要	282
16.9.5 错误6：使用更多的硬件解决所有的性能问题	282
16.9.6 错误5：使用操作型键连接维度和事实	282
16.9.7 错误4：忽视对事实粒度的声明并混淆事实粒度	282
16.9.8 错误3：使用报表设计维度模型	283
16.9.9 错误2：希望用户查询规范化的原子数据	283
16.9.10 错误1：违反事实和维度的一致性要求	283
16.10 本章小结	284
第17章 Kimball DW/BI生命周期概述	285
17.1 生命周期路标	286
17.2 生命周期初始活动	287
17.2.1 程序/项目规划与管理	287
17.2.2 业务需求定义	290
17.3 生命周期技术路径	294
17.3.1 技术架构设计	294
17.3.2 产品选择与安装	296
17.4 生命周期数据路径	297
17.4.1 维度建模	297
17.4.2 物理设计	297
17.4.3 ETL设计与开发	299
17.5 生命周期BI应用路径	299
17.5.1 BI应用规范	299
17.5.2 BI应用开发	299
17.6 生命周期总结活动	300
17.6.1 部署	300
17.6.2 维护和发展	300
17.7 应当避免的常见错误	301
17.8 本章小结	302
第18章 维度建模过程与任务	303
18.1 建模过程概述	303
18.2 组织工作	304
18.2.1 确定参与人，特别是业务代表们	304

18.2.2	业务需求评审	305
18.2.3	利用建模工具	305
18.2.4	利用数据分析工具	306
18.2.5	利用或建立命名规则	306
18.2.6	日历和设施的协调	306
18.3	维度模型设计	307
18.3.1	统一对高层气泡图的理解	307
18.3.2	开发详细的维度模型	308
18.3.3	模型评审与验证	311
18.3.4	形成设计文档	312
18.4	本章小结	312
第19章	ETL子系统与技术	313
19.1	需求综合	314
19.1.1	业务需求	314
19.1.2	合规性	314
19.1.3	数据质量	314
19.1.4	安全性	315
19.1.5	数据集成	315
19.1.6	数据延迟	316
19.1.7	归档与世系	316
19.1.8	BI发布接口	316
19.1.9	可用的技能	317
19.1.10	传统的许可证书	317
19.2	ETL的34个子系统	317
19.3	获取：将数据插入到数据仓库中	318
19.3.1	子系统1：数据分析	318
19.3.2	子系统2：变化数据获取系统	319
19.3.3	子系统3：获取系统	320
19.4	清洗与整合数据	321
19.4.1	提高数据质量文化与过程	322
19.4.2	子系统4：数据清洗系统	323
19.4.3	子系统5：错误事件模式	324
19.4.4	子系统6：审计维度装配器	325
19.4.5	子系统7：重复数据删除(deduplication)系统	326
19.4.6	子系统8：一致性系统	326
19.5	发布：准备展现	328
19.5.1	子系统9：缓慢变化维度管理器	328
19.5.2	子系统10：代理键产生器	332
19.5.3	子系统11：层次管理器	332
19.5.4	子系统12：特定维度管理器	333
19.5.5	子系统13：事实表建立器	335
19.5.6	子系统14：代理键流水线	336
19.5.7	子系统15：多值维度桥接表建立器	337
19.5.8	子系统16：迟到数据处理器	338
19.5.9	子系统17：维度管理器系统	339
19.5.10	子系统18：事实提供者系统	339
19.5.11	子系统19：聚集建立器	340
19.5.12	子系统20：OLAP多维数据库建立器	340
19.5.13	子系统21：数据传播管理器	340
19.6	管理ETL环境	341
19.6.1	子系统22：任务调度器	341
19.6.2	子系统23：备份系统	342
19.6.3	子系统24：恢复与重启系统	343
19.6.4	子系统25：版本控制系统	344

19.6.5 子系统26: 版本迁移系统	345
19.6.6 子系统27: workflow监视器	345
19.6.7 子系统28: 排序系统	346
19.6.8 子系统29: 世系及依赖分析器	346
19.6.9 子系统30: 问题提升系统	346
19.6.10 子系统31: 并行/流水线系统	347
19.6.11 子系统32: 安全系统	347
19.6.12 子系统33: 合规性管理器	348
19.6.13 子系统34: 元数据存储库管理器	350
19.7 本章小结	350
第20章 ETL系统设计与开发过程和任务	351
20.1 ETL过程概览	351
20.2 ETL开发规划	351
20.2.1 第1步: 设计高层规划	352
20.2.2 第2步: 选择ETL工具	352
20.2.3 第3步: 开发默认策略	353
20.2.4 第4步: 按照目标表钻取数据	354
20.2.5 开发ETL规范文档	355
20.3 开发一次性的历史加载过程	356
20.3.1 第5步: 用历史数据填充维度表	356
20.3.2 第6步: 完成事实表历史加载	360
20.4 开发增量式ETL过程	363
20.4.1 第7步: 维度表增量处理过程	363
20.4.2 第8步: 事实表增量处理过程	365
20.4.3 第9步: 聚集表与OLAP加载	367
20.4.4 第10步: ETL系统操作与自动化	368
20.5 实时的影响	368
20.5.1 实时分类	369
20.5.2 实时结构权衡	370
20.5.3 展现服务器上的实时分区	371
20.6 本章小结	372
第21章 大数据分析	373
21.1 大数据概览	373
21.1.1 扩展的RDBMS结构	374
21.1.2 MapReduce/Hadoop结构	375
21.1.3 大数据结构比较	376
21.2 推荐的应用于大数据的最佳实践	376
21.2.1 面向大数据管理的最佳实践	376
21.2.2 面向大数据结构的最佳实践	377
21.2.3 应用于大数据的数据建模最佳实践	381
21.2.4 大数据的数据治理最佳实践	383
21.3 本章小结	384
• • • • •	(收起)

[数据仓库工具箱（第3版）_下载链接1](#)

标签

数据仓库

数据仓库建模

大数据

BI

数据库

计算机

数据挖掘

数据

评论

谢天谢地再版了，前两版市面上都绝迹了，二手都买不到。之前只能硬嚼网上找来的原版书，幸而用Onenote做笔记摘录的话，PDF还比较方便。不过第二版终究是10多年前的，怕是不够与时俱进。

为了实习而看的 睡前催眠读物

看到第六章，然后彻底被翻译弄崩溃了。我要去图书馆找第二版看。

翻译得读起来要崩溃，原书的目录章节也太多太碎，知识点太零散。

kimball维度建模第三版，英文初版1996年，二版1998年，三版2015年。kimball是本书作者，也是kimball集团的创始人。第三版共21章其中14章是14个行业的建模案例。

好书，写的很详细，适合有一定维度建模经验的人带着思考去读，会体会到很多东西

算是数仓的权威了吧，kimball这个名字有点可爱

数仓最流行的理论基础，每个数仓从业者都应该看一下。但是，书里的案例都比较偏传统数仓，很多做法在互联网数仓里已经不适用了。案例部分，我更推崇的还是《大数据之路》里阿里的案例。再吐槽下，这翻译真是渣。

读过的教材中翻译得最烂的一本。

维度建模的经典之作，但是太理论，外国人写的书比较佶屈聱牙，只能不求甚解的去读了

很有价值的一本书，行业问题描述都是我碰到过的，简直是直接指明解决方向。很多地方还没消化好，需要多刷。 PS:中文版比原版难懂。

技术类工具参考书，暂不读。

数据仓库设计全流程，对于我来说，只是走马观花地过了一遍大略。有一些值得注意的点“要控制住规范化的欲望，因为得不偿失”“维度建模要基于最细粒度的事实”

数据仓库发展至今，由于存储成本的大幅降低，冗余的维度信息大量存在于事实表中，已使事实表不再满足第三范式，大大提高了查询效率。
由于维度和事实可以相互转化，实际上任何事实甚至无事实都可以被加工成维度。

翻译有点烂。。。

翻译太差

需要进一步细读

强烈建议看英文，翻译的真的非常之无语！

这本书可以称得上是数据仓库维度建模领域的圣经了

神书

1、选取要建模的业务处理过程，了解整个业务处理过程而不是根据业务部门设计模型
2、定义业务处理的粒度 3、选定用于每个事实表行的维度
4、确定用于形成每个事实表行的数字型事实
维度建模在实施时必须是：业务用户需求+源数据本身
缓慢变化纬度讲的也挺好，就是俗称的拉链表

[数据仓库工具箱（第3版）_下载链接1](#)

书评

经典数仓架构书，翻译非常烂，影响到对于内容的理解。我需要找到第二版看一下，或者直接看英文原版，这个完全看不下去。。。。我带着多值维度的问题去翻书：
如何处理多值维度？如求职者的职业技能、病人的诊断和症状
一般来说，有两个选择：位置设计和桥接表设计 位置设计...

看中译版，翻译质量太差，每个字都认识但连起来就是读不通，感觉语文不够用了

读了前3章，我能看懂翻译的句子就不错了，前面的理论性太多，读完对各个理论没有特别清晰的认识，直接从案例开始了，但愿案例能说的明白点。凑字，凑字，凑字，凑字，凑字，凑字，凑字，凑字，凑字，凑字，凑字，凑字，凑字，凑字，凑字，凑字，凑字，凑字，凑字，

年后上班就在亚马逊下定了此书，结果亚马逊要15后才发货，收到后周末刚拜读了前两章，感觉翻译极差，语句晦涩难懂，没看过前两版，不知好不好，但建议有能力的，还是看英文原版吧。 凑个字 凑个字 凑个字

原版力荐，中文版你懂的
这么重要的数据仓库建设方面的内容，有很多的数据仓库建模方面的知识，就被谭老师翻译成这样了，让国人怎么看啊，google翻译啊有木有？看了两段就看不下去了，专业术语不要直接翻译过来好不好？英文都比中文好懂，无力吐槽。。

原版力荐，中文版你懂的 去掉一个最低分，去掉一个最高分，最后得分0分
必须吐槽，这么重要的数据仓库建设方面的内容，有很多的数据仓库建模方面的知识，就被谭老师翻译成这样了，让国人怎么看啊，google翻译啊有木有？看了两段就看不下去了，专业术语不要直接翻译过来好不好？ ...

原版力荐，中文版你懂的
必须吐槽，这么重要的数据仓库建设方面的内容，有很多的数据仓库建模方面的知识，就被谭老师翻译成这样了，让国人怎么看啊，google翻译啊有木有？看了两段就看不下去了，专业术语不要直接翻译过来好不好？英文都比中文好懂，无力吐槽。。同意楼上意见， ...

正如1楼所说的，书是好书，但中文版翻译的实在太烂了
貌似电子工业出版社已经不在出版了，在taobao上面买了中文拷贝版的，没办法，时间赶+英文不到家，但是，实在郁闷，翻译的连我这种英语不在行的人都能看出来有问题，现在是一边看中文，一边看电子版的E文。

数据仓库建模的必读的好书.有大量实例,便于理解.
但该书的翻译质量实在太差,术语不标准,语句晦涩.建议大家读原版,或者出版社重新找人翻译.

[数据仓库工具箱（第3版）_下载链接1](#)