

大数据管理



[大数据管理_下载链接1](#)

著者:April Reeve

出版者:机械工业出版社

出版时间:2014-3-27

装帧:

isbn:9787111459057

本书是数据集成领域的经典著作，由具有数十年从业经验的资深数据集成专家撰写，数

据管理专家作序推荐！它为大数据时代的大中型企业管理企业内部大量的、复杂的应用系统之间的数据提供了解决方案，全面而深入地讲解数据集成的工具、方法、技巧、解决方案以及最佳实践。

本书分为四部分，共22章，高屋建瓴地阐述了在大型组织环境中，不同计算机系统之间传输数据，以及将异构数据进行集成所用到的技巧、技术和最佳实践，内容涵盖数据集成导论、批处理数据集成、实时数据集成和大数据集成等。

本书虽然介绍了各种数据集成问题的多种不同类型的技术解决方案，但读者无需具备广阔的技术背景就能理解，适合数据处理相关的项目经理、数据分析师、数据模型设计师、数据库工作者以及数据集成程序员等相关技术人员及数据管理专业学生阅读。

作者介绍:

April

Reeve，资深数据集成专家、资深信息管理顾问、企业架构师和项目经理，拥有数十年行业从业经验，经历丰富，曾服务于多家大型跨国公司，主要负责开发数据战略以及管理开发和运行方案。她是数据转换、数据仓库、商务智能、主数据管理、数据集成和数据治理等数据管理领域的专家。目前，她作为企业信息管理实践的咨询顾问服务于EMC 2咨询公司。

目录: 译者序

序言

前言

第一部分 数据集成导论

第1章 数据集成的重要性 / 2

1.1 数据接口的天然复杂性 / 2

1.2 购买供应商应用包的数量日益增加 / 3

1.3 大数据和虚拟化的催化剂 / 3

第2章 什么是数据集成 / 5

2.1 运动中的数据 / 5

2.2 集成为通用格式—数据转换 / 5

2.3 数据从一个系统迁移到另一个系统 / 6

2.4 在组织内部移动数据 / 6

2.5 从非结构化数据中抽取信息 / 8

2.6 将处理移动到数据端 / 9

第3章 数据集成的类型和复杂性 / 10

3.1 管理运动中的数据和持久化数据的异同点 / 10

3.2 批处理数据集成 / 10

3.3 实时数据集成 / 11

3.4 大数据集成 / 11

3.5 数据虚拟化 / 12

第4章 数据集成开发过程 / 13

4.1 数据集成开发生命周期 / 13

4.2 包含业务知识和专家经验 / 14

第二部分 批处理数据集成

第5章 批处理数据集成简介 / 18

5.1 什么是批处理数据集成 / 18

5.2 批处理数据集成生命周期 / 19

第6章 抽取、转换和加载 / 20

6.1 什么是ETL / 20

6.2 概要分析 / 20

6.3 抽取 / 21
6.4 暂存 / 22
6.5 访问层次 / 22
6.6 转换 / 23
6.6.1 简单映射 / 23
6.6.2 查找表 / 24
6.6.3 聚合和规范化 / 24
6.6.4 计算 / 24
6.7 加载 / 24
第7章 数据仓库 / 26
7.1 什么是数据仓库 / 26
7.2 企业数据仓库架构中的层次 / 26
7.2.1 操作型应用层 / 26
7.2.2 外部数据 / 27
7.2.3 数据仓库中的数据暂存区 / 27
7.2.4 数据仓库数据结构 / 28
7.2.5 从数据仓库到数据集市或者商务智能层的暂存 / 28
7.2.6 商务智能层 / 28
7.3 加载到数据仓库中的数据类型 / 29
7.3.1 数据仓库中的主数据 / 29
7.3.2 数据仓库中的余额和快照数据 / 30
7.3.3 数据仓库中的事务型数据 / 31
7.3.4 事件 / 31
7.3.5 调整 / 31
第8章 数据转换 / 39
8.1 什么是数据转换 / 39
8.2 数据转换生命周期 / 39
8.3 数据转换分析 / 39
8.4 数据加载最佳实践 / 40
8.5 提高源数据质量 / 40
8.6 映射到目标系统 / 41
8.7 配置数据 / 41
8.8 测试和依赖 / 42
8.9 私有数据 / 42
8.10 校对 / 43
8.11 环境 / 43
第9章 数据归档 / 47
9.1 什么是数据归档 / 47
9.2 归档数据选择 / 47
9.3 已归档数据可以恢复吗 / 48
9.4 归档环境下数据结构的确认 / 48
9.5 灵活的数据结构 / 49
第10章 批处理数据集成架构和元数据 / 54
10.1 什么是批处理数据集成架构 / 54
10.2 概要分析工具 / 55
10.3 建模工具 / 55
10.4 元数据存储库 / 55
10.5 数据移动 / 56
10.6 转换 / 56
10.7 调度 / 57
第三部分 实时数据集成
第11章 实时数据集成简介 / 64
11.1 为什么需要实时数据集成 / 64
11.2 为什么需要两组技术 / 64

第12章 数据集成模式 / 66	
12.1 交互模式 / 66	
12.2 松耦合 / 66	
12.3 中心和节点模式 / 66	
12.4 同步交互和异步交互 / 69	
12.5 请求和应答 / 70	
12.6 发布和订阅 / 70	
12.7 两阶段提交 / 70	
12.8 集成交互类型 / 71	
第13章 核心实时数据集成技术 / 72	
13.1 令人困惑的术语 / 72	
13.2 企业服务总线 / 72	
13.3 面向服务架构 / 75	
13.4 可扩展标记语言 / 77	
13.5 数据复制和变化数据捕获 / 81	
13.6 企业应用集成 / 82	
13.7 企业信息集成 / 82	
第14章 数据集成建模 / 84	
14.1 规范化建模 / 84	
14.2 消息建模 / 88	
第15章 主数据管理 / 89	
15.1 主数据管理简介 / 89	
15.2 需要主数据管理方案的原因 / 89	
15.3 购买的软件包与主数据 / 90	
15.4 参考数据 / 90	
15.5 主和从 / 91	
15.6 外部数据 / 93	
15.7 主数据管理功能 / 93	
15.8 主数据管理方案的类型—注册表以及数据中心 / 94	
第16章 实时更新数据仓库 / 95	
16.1 企业信息工厂 / 95	
16.2 操作型数据存储 / 96	
16.3 移动到数据仓库的主数据 / 97	
第17章 实时数据集成架构和元数据 / 99	
17.1 实时数据集成元数据简介 / 99	
17.2 建模 / 100	
17.3 概要分析 / 100	
17.4 元数据库 / 101	
17.5 企业服务总线—数据转换和调度 / 101	
17.5.1 技术中介 / 101	
17.5.2 业务内容 / 102	
17.6 数据移动和中间件 / 102	
17.7 外部交互 / 102	
第四部分 大数据集成	
第18章 大数据集成简介 / 106	
18.1 数据集成及非结构化数据 / 106	
18.2 大数据、云数据及数据虚拟化 / 106	
第19章 云架构和数据集成 / 107	
19.1 为什么云中的数据集成比较重要 / 107	
19.2 公共云 / 107	
19.3 云安全 / 108	
19.4 云延迟 / 109	
19.5 云冗余 / 110	
第20章 数据虚拟化 / 111	

20.1 恰逢其时的一项技术 /	111
20.2 数据虚拟化的商业用途 /	112
20.2.1 商务智能方案 /	112
20.2.2 集成不同类型的数据 /	113
20.2.3 快速向数据仓库中增加或者原型增加数据 /	113
20.2.4 将物理上不同的数据一起展现 /	113
20.2.5 利用不同的数据和模型触发交易 /	114
20.3 数据虚拟化架构 /	114
20.3.1 源和适配器 /	114
20.3.2 映射、模型和视图 /	114
20.3.3 转换和展现 /	115
第21章 大数据集成 /	116
21.1 什么是大数据 /	116
21.2 大数据维度—量 /	116
21.2.1 大规模并行处理—将处理过程移动到数据端 /	116
21.2.2 Hadoop和MapReduce /	117
21.2.3 与外部数据集成 /	117
21.2.4 虚拟化 /	118
21.3 大数据维度—多样性 /	118
21.3.1 数据类型 /	118
21.3.2 集成不同类型的数据 /	118
21.4 大数据维度—速度 /	120
21.4.1 流式数据 /	121
21.4.2 传感器和GPS数据 /	121
21.4.3 社会化媒体数据 /	121
21.5 传统大数据应用案例 /	121
21.6 更多大数据应用案例 /	122
21.6.1 医疗 /	122
21.6.2 物流 /	122
21.6.3 国家安全 /	122
21.7 利用大数据的力量—实施决策支持 /	123
21.7.1 触发行动 /	123
21.7.2 从内存以及磁盘中检索数据的速度 /	123
21.7.3 从数据分析到模型，从流式数据到决策 /	124
21.8 大数据架构 /	125
21.8.1 操作型系统和数据存储 /	125
21.8.2 中间数据中心 /	126
21.8.3 商务智能工具 /	126
21.8.4 数据虚拟化服务器 /	127
21.8.5 批处理和实时数据集成工具 /	127
21.8.6 分析型沙盒 /	127
21.8.7 风险响应系统/推荐引擎 /	127
第22章 移动数据管理总结 /	132
22.1 数据集成架构 /	132
22.1.1 为什么需要数据集成架构 /	132
22.1.2 数据集成生命周期和专家经验 /	132
22.1.3 安全和隐私 /	133
22.2 数据集成引擎 /	134
22.2.1 操作连贯性 /	134
22.2.2 ETL引擎 /	134
22.2.3 企业服务总线 /	135
22.2.4 数据虚拟化服务器 /	135
22.2.5 数据移动 /	136
22.3 数据集成中心 /	136

- 22.3.1 主数据 / 137
- 22.3.2 数据仓库和操作型数据存储 / 137
- 22.3.3 企业内容管理 / 138
- 22.3.4 数据归档 / 138
- 22.4 元数据管理 / 138
 - 22.4.1 数据发现 / 138
 - 22.4.2 数据概要分析 / 139
 - 22.4.3 数据建模 / 139
 - 22.4.4 数据流建模 / 139
 - 22.4.5 元数据存储库 / 139
- 22.5 结束语 / 140
- 参考文献 / 141
- (收起)

[大数据管理 下载链接1](#)

标签

大数据

数据管理

技术

理论

集成

法学与经济

数据分析

数据仓库

评论

概括全面的介绍大数据管理技术，原有技术如何应用在大数据，和新的大数据技术的应用。没有太过于深入的东西。对各种技术在大数据挑战下应用和实施的访谈挺有趣，可以了解多些在考虑大数据解决方案时的考虑和实际的挑战。

连教材都算不上，基本上可以算是一个很枯燥的博士论文，十分生硬，既没有实例也没有解释说明或深刻见解。完全是一堆概念堆积起来的一本书

好吧，图书馆借的书，翻了翻，除了一些名词和解释，没有任何有价值的内容。

本书前半部分对于数据集成的概念、逻辑、技术进行了较为清晰的介绍和讲解，后半篇大多是概念性的描述和介绍，没有深入，浅尝辄止，读起来没有滋味

[大数据管理 下载链接1](#)

书评

[大数据管理 下载链接1](#)