

数据仓库应用指南



[数据仓库应用指南_下载链接1](#)

著者:Robert Laberge

出版者:机械工业出版社华章公司

出版时间:2012-3-24

装帧:

isbn:9787111370444

本书全面系统地讲解如何规划、设计、构建和管理数据仓库/商务智能解决方案。介绍

在数据仓库开发项目中如何激励用户，在整个企业范围内更好地驱动决策制定，从专业的开发和培训人员获取详细的指导和最佳实践经验。本书内容涉及如何选择恰当的组件、构建企业数据模型、配置数据集市和数据仓库、构建数据流并降低风险，还涉及项目开发中变更管理、数据监理和安全方面的问题。

主要内容：

- 理解BI和数据仓库系统组件
- 建立项目目标并有效制定部署计划
- 用数据挖掘发现组织内业务见解
- 使用ETL技术对数据进行输入、清洗和规范化
- 使用结构化输入文件来定义数据需求
- 使用自上而下、自下而上、混合的设计方式
- 使用数据监理工具来处理安全和性能优化问题

作者介绍:

Robert (Bob)

Laberge是多家互联网企业的创始人、IBM产业模式和资产实验室的首席顾问，他的研究重点是数据仓库和商务智能解决方案。

Bob早在20世纪70年代末就开始其职业生涯，当时比尔·盖茨还只是一个百万富翁，Bob曾经是开发人员、数据库管理员、数据建模师、项目经理、数据架构师、企业信息架构师、数据仓库/商务智能审计员、战略师，而且还是富于创新的企业家。从那时，Bob就跑遍全球，通过设计、优化、最佳实践和在概念层、逻辑层和物理层的常识说明，提供指导、培训和证明数据仓库和商务智能实践经验和解决方案。Bob成功地帮助了50多家大型企业扩展业务，这些企业涉及零售、保险、医疗、铁路、电信、电子商务和银行等行业。

Bob拥有英国Durham大学的工商管理硕士学位。你可以通过datawarehousementor@gmail.com联系他。

目录: 译者序

前言

作者简介

第一部分 准备

第1章 数据仓库和商务智能概述

1.1 商务智能概述

1.1.1 定义

1.1.2 商务智能的价值

1.1.3 剖析商务智能

1.1.4 商务智能的成功要素

1.1.5 商务智能的目标

1.1.6 BI用户展现层

1.1.7 BI工具和架构

1.1.8 全球化带来的发展

- 1.2 数据仓库概述
 - 1.2.1 定义
 - 1.2.2 数据仓库系统
 - 1.2.3 数据仓库架构
 - 1.2.4 数据流术语
 - 1.2.5 数据仓库目标
 - 1.2.6 数据结构化策略
 - 1.2.7 数据仓库业务
- 1.3 常见问题
 - 1.3.1 当前系统是否足够好
 - 1.3.2 数据仓库的价值
 - 1.3.3 成本多高
 - 1.3.4 时间多长
 - 1.3.5 成功的因素
- 第2章 企业中的数据
 - 2.1 企业资产
 - 2.1.1 具有上下文的数据
 - 2.1.2 数据质量
 - 2.1.3 数据字典
 - 2.1.4 数据组件
 - 2.2 组织数据
 - 2.2.1 对数据结构化
 - 2.2.2 数据模型
 - 2.2.3 数据架构
 - 2.3 竞争优势
 - 2.3.1 构建还是购买数据模型
 - 2.3.2 指导业务
- 第3章 为什么创建数据仓库
 - 3.1 平台迁移
 - 3.1.1 业务连续性
 - 3.1.2 逆向工程
 - 3.1.3 数据质量
 - 3.1.4 并行环境
 - 3.1.5 附加值
 - 3.2 数据仓库集中化
 - 3.2.1 企业间并购
 - 3.2.2 企业内合并
 - 3.2.3 集中式设计和局部使用
 - 3.3 数据集市整合
 - 3.4 新方案
 - 3.5 新方案：动态报表
 - 3.6 “Just Build It” 模式
 - 3.7 数据Foundation
 - 3.8 不构建数据仓库的原因
 - 3.8.1 数据质量差
 - 3.8.2 缺乏商业目标
 - 3.8.3 缺乏管理层支持
 - 3.8.4 目标不明确
 - 3.8.5 当前系统足够用
 - 3.8.6 缺乏人才资源
 - 3.8.7 环境不稳定
 - 3.8.8 成本太高
 - 3.8.9 管理不善
- 第4章 数据仓库和商务智能战略

- 4.1 商务智能战略
 - 4.1.1 商业目标
 - 4.1.2 商业用途
 - 4.1.3 架构概览
- 4.2 数据仓库战略
 - 4.2.1 用途
 - 4.2.2 数据仓库架构
- 4.3 重点和成功
 - 4.3.1 整个企业还是业务线
 - 4.3.2 目标明确
 - 4.3.3 成功：衡量的标准是什么
- 4.4 从何处着手
 - 4.4.1 关于商务智能
 - 4.4.2 关于数据仓库
- 4.5 如何开始
 - 4.5.1 关于商务智能
 - 4.5.2 关于数据仓库
- 4.6 项目阶段化
- 4.7 需要多长时间（重新回顾）
- 4.8 兴趣点
 - 4.8.1 常见的失败原因
 - 4.8.2 基本原则
- 第5章 项目资源：角色和洞察力
 - 5.1 关键点
 - 5.1.1 项目团队
 - 5.1.2 资深专业知识
 - 5.1.3 领导力
 - 5.1.4 项目发起人
 - 5.1.5 数据仓库管理层
 - 5.2 团队结构
 - 5.2.1 管理层发起人
 - 5.2.2 数据管家
 - 5.2.3 基本资源
 - 5.3 定期审查：进度审核
 - 5.4 能力中心
- 第6章 项目总结概论
 - 6.1 项目章程
 - 6.2 项目范畴
 - 6.3 工作说明书
- 第二部分 组件
- 第7章 商务智能：数据集市及其使用方式
 - 7.1 为什么要对数据建模
 - 7.1.1 数据模型的类型
 - 7.1.2 数据设计
 - 7.2 事实表
 - 7.2.1 事实的类型
 - 7.2.2 事实表的类型
 - 7.2.3 衡量指标来源
 - 7.2.4 事实表关键字
 - 7.2.5 事实表粒度
 - 7.2.6 事实表密度
 - 7.2.7 无事实的事实表
 - 7.3 维度表
 - 7.3.1 维度还是指标

- 7.3.2 历史表和日期表
- 7.3.3 维度表关键字
- 7.3.4 维度表的粒度
- 7.3.5 维度属性的来源和价值
- 7.3.6 维度类型
- 7.3.7 级别和辅助表
- 7.3.8 个人信息表
- 7.3.9 维度数
- 7.4 规模
- 第8章 企业数据模型
- 8.1 数据模型概览
- 8.2 构建企业数据模型的目标
- 8.3 企业数据模型的好处
- 8.4 数据模型：从何处开始
- 8.5 完全自上而下的数据模型
- 8.5.1 主题领域模型
- 8.5.2 概念模型
- 8.5.3 实体关系模型
- 8.6 总线结构
- 8.7 购买的数据模型
- 8.8 模型分析
- 8.8.1 数据组件
- 8.8.2 范化数据模型
- 8.8.3 超类和子类模型
- 8.8.4 在范化的数据模型中收集历史信息
- 8.8.5 代理键
- 8.8.6 逻辑和物理数据模型
- 8.8.7 是否具备参照完整性
- 8.9 其他数据模型
- 8.9.1 输入数据模型
- 8.9.2 临时存储数据模型
- 8.10 最后的思考
- 第9章 数据仓库架构：组件
- 9.1 架构概述
- 9.2 架构师角色
- 9.2.1 解决方案架构师
- 9.2.2 数据仓库架构师
- 9.2.3 技术架构师
- 9.2.4 数据架构师
- 9.2.5 ETL架构师
- 9.2.6 BI架构师
- 9.2.7 综合
- 9.3 体系结构分层
- 9.3.1 单层体系结构
- 9.3.2 经典的两层体系结构
- 9.3.3 高级的三层体系结构
- 9.4 数据仓库架构
- 9.4.1 单独的数据集市架构
- 9.4.2 总线结构
- 9.4.3 中央存储库架构
- 9.4.4 联合架构
- 9.5 组件（分层）
- 9.5.1 数据源
- 9.5.2 数据生成

- 9.5.3 数据组织
- 9.5.4 数据分发
- 9.5.5 信息输出
- 9.6 实现方式
 - 9.6.1 数据设计和数据流
 - 9.6.2 逻辑和物理模型
 - 9.6.3 自上而下的方式
 - 9.6.4 自下而上的方式
 - 9.6.5 混合模式
- 9.7 捷径
 - 9.7.1 数据采集层
 - 9.7.2 中央数据层
 - 9.7.3 数据分发层
 - 9.7.4 表现层
 - 9.7.5 用户展现层
 - 9.7.6 方法论
 - 9.7.7 现成的解决方案
- 第10章 ETL和数据质量
 - 10.1 架构
 - 10.1.1 数据获取
 - 10.1.2 数据分发
 - 10.1.3 ETL映射
 - 10.1.4 初始加载和增量加载
 - 10.1.5 ETL、ELT和ETTL
 - 10.1.6 并行操作
 - 10.1.7 ETL功能角色
 - 10.1.8 数据流图
 - 10.1.9 业务数据存储系统
 - 10.2 数据源系统
 - 10.2.1 没有数据源
 - 10.2.2 多个数据源
 - 10.2.3 其他来源（结构化输入文件）
 - 10.2.4 非结构化数据
 - 10.3 数据剖析
 - 10.4 数据获取
 - 10.4.1 多个大文件
 - 10.4.2 伪文件
 - 10.4.3 故障预防策略
 - 10.5 转换和临时数据存储
 - 10.5.1 准备工作
 - 10.5.2 代理键
 - 10.5.3 参照完整性
 - 10.5.4 聚合、分析和汇总
 - 10.5.5 编码表
 - 10.6 加载
 - 10.6.1 是否加载历史数据
 - 10.6.2 插入、更新、插入或更新、删除
 - 10.6.3 数据获取信息
 - 10.6.4 加载调度
 - 10.7 企业数据仓库的临时数据存储和总线架构的临时数据存储
 - 10.8 数据分发
 - 10.9 数据质量
 - 10.10 ETL工具
- 第11章 项目规划和方法论

- 11.1 基础
 - 11.1.1 风险：逐步发展
 - 11.1.2 风险：数据质量
 - 11.1.3 风险：资源
 - 11.1.4 风险：成本
 - 11.1.5 变更管理
 - 11.1.6 最佳实践
- 11.2 错误
- 11.3 项目规划方法论
 - 11.3.1 业务需求分析
 - 11.3.2 战略和规划
 - 11.3.3 解决方案纲要
 - 11.3.4 设计
 - 11.3.5 构建
 - 11.3.6 部署
 - 11.3.7 使用
- 第三部分 构建
- 第12章 工作场景
 - 12.1 让我们开始“烹饪”吧
 - 12.2 自上而下
 - 12.2.1 字典
 - 12.2.2 集中式数据模型
 - 12.2.3 数据架构
 - 12.2.4 数据源
 - 12.2.5 数据模型
 - 12.2.6 数据库
 - 12.2.7 数据获取
 - 12.2.8 解决方案概述
 - 12.3 自下而上
 - 12.3.1 最终结果
 - 12.3.2 字典
 - 12.3.3 数据架构
 - 12.3.4 一致性维度的管理
 - 12.3.5 数据源
 - 12.3.6 解决方案概述
 - 12.4 混合式
 - 12.4.1 起步工作
 - 12.4.2 数据模型
 - 12.4.3 数据架构
 - 12.4.4 解决方案概述
 - 12.5 归并
 - 12.6 没有输入：结构化的输入文件
 - 12.7 集成的第二阶段
 - 12.8 更大的框架：企业信息架构
- 第13章 数据监理
 - 13.1 什么是数据监理
 - 13.2 数据监理的原因
 - 13.3 企业结构
 - 13.4 驱动和启动
 - 13.5 数据监理的主要方面
 - 13.5.1 安全性和敏感性
 - 13.5.2 数据质量
 - 13.5.3 所有权
 - 13.5.4 变更控制

13.6 数据监理的准备工作

第14章 项目后评审

14.1 概述

14.2 项目评审

14.3 后续工作

• • • • • (收起)

[数据仓库应用指南_下载链接1](#)

标签

数据仓库

商业智能

项目

系统开发

管理

科技大类

数据建模

数据库

评论

例子深度恰到好处，数据模型和整体构架讲解透彻，对两大流派的比较也相当中肯。美中不足的是重复内容比较多，虽然一些内容在提高深度，但更多的真的是在炒冷饭了。此外，本书并没有实践课，最多算是课上例题吧。想知道DW/BI系统的开发流程和整体框架，就是这本书了。本书最赞之处是把数据仓库和商业智能结合起来讲解，数据仓库是手段，商业智能是目的。

我认为，不论是被《数据仓库》启蒙还是从《数据仓库工具箱》出发，一定都有过疑问，哪一派的招式更强，是否另外一种架构思路更好使。这本书的最大亮点即是把Inmon和Kimball的理论(总线架构/中央存储架构)的优劣分析得十分清楚。并提出了"联合架构"的方案(可操作性另说)。

1.在高层面上介绍了数据仓库，以及数据仓库项目管理实施的最佳实践。不涉及太多技术细节，适合高管扫盲，全员培训，以及程序员入门。

对于新手来说，这本还是比数据仓库那本容易理解的多的…

听说的：这本书勉强强算是可以，内容太泛泛而谈，空洞无物，对于开发人员而言不是很实用

又是大学教材般的封面，极烂。这本我只花了11分钟就翻完了，就看了几张图。

[数据仓库应用指南_下载链接1](#)

书评

忍不住写这篇书评，是看到相关售书网站对此书评价有些低。我来中和一下影响，免得好书被埋没。如果你寻找的是一份“数据仓库建设How to”或者“数据仓库建设step by step”，那么这本书是不合适的。它突出的是过程顾问，又尽可能的剔除掉了许多技术边角料，几乎提纲挈领…

[数据仓库应用指南_下载链接1](#)