



随书附赠CD-ROM

全球最畅销的数据仓库项目管理指南，
完全源于实战专家的经验教训



Data Warehouse Project Management

数据仓库项目管理

[美] 锡德·阿德尔曼 拉里萨·特佩卢克·莫斯 /著 薛宇 王剑锋 /译
(Sid Adelman & Larissa Terpeluk Moss)



清华大学出版社



1
软件管理与软件工程译丛

数据仓库项目管理

[美] 锡德·阿德尔曼 拉里萨·特佩卢克·莫斯 / 著 薛宇 王剑锋 / 译
(Sid Adelman & Larissa Terpeluk Moss)



清华大学出版社

Data Warehouse Project Management

Sid Adelman, Larissa Terpeluk Moss

EISBN 0-201-61635-1

Copyright © 2000 by Sid Adelman and Larissa Terpeluk Moss.

Original English language edition published by Addison-Wesley Longman, Inc.

All rights reserved.

本书中文简体字版由 Pearson Education 授权清华大学出版社在中国境内(香港、澳门特别行政区和台湾地区除外)独家出版、发行。

未经出版者书面许可,不得以任何方式复制或抄袭本书的任何部分。

本书封面贴有 Pearson Education 防伪标签,无标签者不得销售。

北京市版权局著作权合同登记号:图字:01 2002-5153

书 名: 数据仓库项目管理

原 著 者: [美] 锡德·阿德尔曼(Sid Adelman)

拉里萨·特佩卢克·莫斯(Larissa Terpeluk Moss)

译 者: 薛 宇 王剑锋

出 版 者: 清华大学出版社(北京清华大学学研大厦,邮编 100084)

<http://www.tup.tsinghua.edu.cn>

责任编辑: 熊妍妍

封面设计: 孙剑波

版式设计: 肖 米

印 刷 者: 清华大学印刷厂

发 行 者: 新华书店总店北京发行所

开 本: 880×1230 1/32 印张: 14.25 插页: 2 字数: 302 千字

版 次: 2003 年 2 月第 1 版 2003 年 2 月第 1 次印刷

书 号: ISBN 7-900643-87-7

印 数: 0001~5000

定 价: 39.00 元(含光盘)

《软件管理与软件工程译丛》

编委会名单

主编

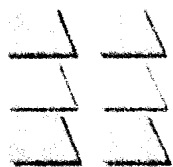
王文京 用友软件股份有限公司, 董事长

总策划

麻众志 闻 洁

编委会委员 (以下按姓氏笔画排序)

万起光	北京湘计立德信息技术有限公司
方春旭	UMLChina
邓俊辉	清华大学计算机系计算机软件研究所
乐林峰	UMLChina
任 伟	北京翰联世纪企业管理技术有限公司
朱于军	朗讯科技(中国)有限公司贝尔实验室
杨作兴	北京方舟科技有限公司
汪 颖	IBM(中国)有限公司上海分公司
陈江北	用友软件股份有限公司
周浩宇	安捷伦科技软件有限公司
孟迎霞	《程序员》杂志社
钱 岭	朗讯科技(中国)有限公司贝尔实验室
熊 节	《程序员》杂志社
潘加宇	UMLChina



策划人语

印象里,从谭浩强老师的《BASIC 语言》在全国畅销,最终销量超过1 200万册时,民族软件产业就承载了国人太多的期待。计算机图书出版者也与读者结下了不解之缘。人们将软件产业看作中华民族腾飞的一次历史性机遇。的确,软件产业是头脑产业,中华民族最不缺的就是有智慧的人才。几台微机,一本指南,似乎就可以开公司编程赚钱,而且可以出口赚美元。然而,20 多年转瞬即逝,我们听说过很多软件英雄和编程奇才的故事,却没有产生一个有国际影响的软件品牌或有国际竞争力的民族软件企业。就在我们身边,印度的软件产业却已迅速崛起,产值 10 年增长了 10 倍,数十家企业通过了 CMM5 级认证,印度已经成为世界第二大软件出口国。目前,中国经济发展水平可能要领先印度 10 年,但中国软件产业却落后印度将近 10 年,原因何在?

几乎所有的人都认为,软件开发是年轻人的职业。程序员们一边挥着汗水,辛苦地熬夜写代码,一边又对自己

30岁以后的职业发展方向充满惶恐。实际上,我国最缺的是有10年以上经验的软件工程师。金山软件公司总裁雷军曾说过:“在印度,包括在美国,我见到的项目经理都是三四十岁的人,他们‘越老越值钱’,有些人甚至拥有超过20年的行业经验。”为什么中国的程序员总是在不断学习新的开发工具、钻研程序代码,而不能逐步提升自己的视野、思维和经验?

作为出版者,我们一直关注着中国软件产业的发展,也时刻检讨着自己的责任。长期以来,在各大城市书城和计算机专业书店的书架上,很难找到几本软件管理和软件工程的图书,更多的书店里甚至根本没有软件工程图书的分类。同样在印度,软件管理的图书比比皆是,数量远远超过编程指南的图书。在全球最大的网上书店亚马逊网站上,软件管理和软件工程方面的图书也是门类清晰,为数众多。对这种强烈反差长期的漠视,也许正是我们出版者的责任?

于是,在2001年的秋季,我们在一个月的时间内,走访了软件行业中的各大专业媒体、网站、院校和企业,与50多位业界专家进行了交流和沟通。最终确定了引进出版一套国外软件管理和软件工程经典图书的设想。希望借此能将国外成熟先进的软件管理思想和管理经验系统引进到国内,从而推动中国的软件产业的发展,对广大软件工程师的成长和素质提升有所帮助。我们相信,中国软件产业的繁荣和发展,不仅有赖于国家产业政策的支持;还有赖于从业者管理理念的突破和管理思想的普及;更有赖于软件产业中有人文精神的优秀人才的涌现和具有核心竞争力的优秀企业的诞生。

经过将近一年的努力,在业界许多专家的支持下,我



们抱着宁缺毋滥的想法,将国外软件管理和软件工程的经典图书完整地筛选了一遍后,选定了这套丛书,其中的每一本均是相关领域中最经典或最具代表性的作品,有的书曾在海外屡获大奖;有的书已经出版 20 年,却仍然畅销不衰。此外,我们对翻译和编辑质量严格把关,虽然两度延期出版,但让我们欣慰的是,如今终于可以向读者和原著者们提交一份负责任的答卷。

在本系列丛书引进和出版的过程中,我们得到了《IT 经理世界》、《程序员》杂志;UMLChina 网站等机构和众多业界专家学者的真诚理解和大力支持,在此向他们一并致以衷心的感谢。读者如有任何意见和建议,请致 hjvp@sohu.com。

和君创业·闲洁工作室

2002 年 10 月



译者序

在一个全球化与完全开放的世界中,Internet 使技术的发明与传播变得极为容易。记得 6 年前,在中国人民大学信息学院的“国际信息技术”课堂上,陈禹院长说到中国在接受计算机技术的速度方面与国外相差已经不足 1 个月了。他以 Java 为例,在美国刚刚出版 1 个月的新书,国内就会出现相应的中译本。而进入 21 世纪以后,这 1 个月的差距也已经被覆盖全球的 Internet 所消融。在中国,技术层面上的国外知识往往最容易被接受和模仿。数据仓库技术近年来的迅速发展和广泛应用就证明了这一点。译者目前就读的中国人民大学财政金融学院也在金融、税收、证券等领域启动了相关的数据仓库项目。如此广泛而热闹地集中使用某一项新技术正好说明了这一技术的生命力。但是,我们常常忽视了一些更为本质和重要的非技术方面的问题:与新技术相配套的管理思维和决策方式。可以说,这一缺陷使得数据仓库对提高企业管理质量的贡献率大打折扣。经常可以看到,许多公司津津乐道于数据

仓库项目的启动与完工,却绝口不提它们在组织结构、人员配备乃至项目开发方面依然还在沿用老套的格局和定式。

这一现象使得“科学技术是第一生产力”成为空话。而我手边这本《数据仓库项目管理》则为我们带来了一种绝佳的解决方案。在本书中,主要探讨了如何对技术进行管理,着力解决了传统的开发环境、老式的组织结构、过时的开发模式、陈旧的人员配备等一系列制约技术发展的瓶颈,为技术效率的提高开辟了大道。作者 Sid Adelman 和 Larissa Terpeluk Moss 在数据仓库项目管理领域已工作多年,深谙此道,穷数年之功而推出的这本力作,用我们国人的话来形容:“是说到了点子上”。本书提供了一套全新的思维方式和价值取向,接受它无异于接受一场头脑的变迁运动。我想,这场运动或早或迟,在现代化的中国国土上是一定会发生的。

在翻译此书的过程中,恰逢我在中国人民大学金融与财税电子化研究所参与一个数据仓库项目。在项目开发的过程中,我引入了书中的许多策略和方法,取得了极佳的效果。其最大的好处是将以前纷繁芜杂的工作变得条理清晰、有章可循,极大地提高了开发速度与质量。至于本书的优点和特点,在 John A. Zachman 先生所写的序中已充分地体现出来了,这里不再赘述。

本书面向数据仓库的开发者、管理者、设计者、数据管理员、数据库管理员以及在现代数据处理环境中的相关人员。对于计算机专业的本科生和研究生也有重要的参考价值。

最后,我要感谢我的导师——谭荣华教授,他在这一领域的渊博知识给了我极大的指导和启迪,这位令人尊敬



的长者所给予我的每一点帮助和关怀,我都将铭记于心。另外,还要感谢我的朋友和同学——王剑锋、杨颖,他们帮助我翻译了一些困难的章节并整理了所有的附录。其中,王剑锋翻译了第 8、第 9 章;杨颖翻译了第 4、第 5、第 6 章。

译者 薛 宇

中国人民大学财政金融学院

2002 年 4 月



英文版前言

你成为项目管理者已经很多年了,并且成功地实现了许多系统,但数据仓库项目对你而言却是全新的事物。你多年积累起来的知识和经验并不能使你在数据仓库项目开发上一帆风顺。你多年来一直恪守的方法并不能有效地帮助你控制项目中的所有活动。任务不得不重复多次,另外,还不得不实施一些你以前从未考虑过的任务;分配给员工的角色和责任显得不够完整,有时甚至是错误的;你的用户不准备在项目上花费太多时间,你无法确认能从他们那里得到何种需求;你知道你的源文件中存在一些不好的数据,但你却不能准确预测它们对于抽取/转换/加载过程的影响。

也许你正在规划你的第一个数据仓库项目,并且已经听说此类项目有一定的难度。而无论你已经管理过数据仓库项目还是正在规划你的首个数据仓库项目,本书都将帮助你成功地完成任务。在你阅读本书之前,我们将本书的组织结构做一个简要的介绍,为你提供一幅概要的

视图。

本书的目的

做数据仓库最困难的就是你所管理的是一个极具动态性的项目。而数据仓库项目之所以是动态的,原因在于项目需求不再像操作型系统那样可以预先定义好。建立数据仓库的过程常常会导致需求的调整或者发现新的需求。而且,参与这些项目的工作人员颇具天赋但往往缺乏经验。新型技术组件的复杂性和学习曲线通常都会被低估。技术方和业务方的管理者都没有理解数据仓库项目的复杂程度,常常向开发组和项目管理者提出不合理的要求。换句话说,这些项目对于管理也是一个巨大的挑战。

本书的目的在于说明数据仓库项目中将会遇到的典型挑战,并教会项目管理者如何认识项目中的障碍和陷阱。我们列举了一些在实际工作中常见的风险和失败的例子,并提出了避免和防范它们的建议,或者至少应当减轻它们对项目的影响。在每一章的末尾都有一块名为“警示故事”的独立内容,它摘要地描述了我们自己的经历。每章都以一个练习结束,通过做练习可以巩固我们所学到的内容。

本书适合哪些人阅读

如果你正尝试做自己的第一个数据仓库项目,对于管理此类项目的独特方式不太熟悉,那么本书将对你大有裨益。如果你已经管理过数据仓库项目,但并不是很成功,你希望在管理下一个项目时做得更好,那么本书将解释你在项目中曾经遇到的困难,并为避免或减轻困难提出有益的建议。

本书不是基础项目管理的入门教材。相反,本书将着重为有经验的项目管理人员提供指导,他们需要知道数据仓库项目与传统项目的区别,并且希望获得数据仓库领域内专家的帮助。

本书的组织结构

我们写作此书的方式是将每一章作为一个相对独立的部分,因为我们意识到许多项目管理者常常把某一章作为他们的主要参考对象。为了做到这一点,不可避免地会在各章重复出现一些交叉性的内容。

每一章的开始都有一个关于主题的简短列表,接下来是我们在该领域的一些经历,对于其中的雷区将特别指明,然后以一份摘要和一套练习作为结束。部分章节还有附录,附录包括模板、工作表或附加指导。我们鼓励你使用这些模板,它们将帮助你在你的企业中实现过程的标准,并简化你的工作。

我们尽量使用中性的人称来写作此书。但有时当我们不得不使用性别时,我们选择了男性的称呼“他”,我们当然清楚有许多杰出的女性项目管理者,但将性别称呼改成“他/她”将会打乱内容的连续性。所以,我们希望读者能够理解这种简略的人称用法。

本书的大部分内容都涉及到了严肃的主题,使用了比较严肃的语气。然而,为了提高阅读的趣味性,我们插入了一些小笑话,有的俏皮话在政治上是不正确的。我们希望读者不要在意这些插科打诨的东西。

无论你计划从头至尾逐章阅读本书还是仅仅把本书作为参考,我们都建议你从第1章“数据仓库概述”开始,以便于理解我们的知识系统和名词术语。本书涉及的所

有主题都在第 12 章“项目规划”中得到了体现,它将所有章节汇集成一幅完整的视图提供给项目管理者。

各章浏览

第 1 章“数据仓库概述”为你介绍数据仓库世界的梗概。它将传统决策支持系统与数据仓库进行对比并列出了两种环境之间的差异。该章还说明了管理数据仓库项目的困难所在,并解释了作者在这方面的立场和观点。

第 2 章“项目目标”深入讨论了传统决策支持系统的缺陷,并说明了数据仓库的短期和长期目标。

第 3 章“成功要素”讨论了成功的标准,描述了决定一个项目成败的因素。它还提到了关键成功要素——项目成功的必备特征以及如何评价结果。

第 4 章“项目风险”描述了各种数据仓库可能遇到的风险种类。它列出了项目固有的风险以及随之相伴的各种后果,并提出了处理它们的技巧。

第 5 章“使用户满意”强调理解业务的重要性,并检查了所有影响用户或被用户所影响的领域,包括从收集需求到与用户进行交流等各个方面。

第 6 章“成本收益分析”讨论了每个数据仓库项目中分析成本的必要性。它处理各类典型的成本和预期收益,并提供了一个模板以便你能够分析自己项目的成本。

第 7 章“选择工具软件”描述了数据仓库工具的清单,解释了如何使工具适应企业的组织架构,讨论了判断产品需求的步骤以及如何拒绝你不想考虑的软件供应商。

第 8 章“组织与文化要素”介绍了数据仓库项目中开发组成员的角色和责任。它还解释了数据仓库开发组的结构并对职员要素,诸如:招聘和保留、培训和交流进行了

讨论。

第9章“方法论”解释了传统的瀑布式方法不适用于数据仓库的原因。它还描述了各种并行开发途径,比如抽取/转换/加载过程、数据提交过程、元数据收集和导航工作。

第10章“数据模型”介绍了逻辑数据模型的细化目的和用途,数据整合的主要技术,以及物理数据模型中数据库设计的目的。它比较了基于二维E R结构的传统数据库设计模式与基于星型结构的多维数据库设计模式。

第11章“数据质量”定义了数据仓库中数据质量的意义和出现劣质数据的代价。它还定义了源文件中发现的各种不好数据的清单并提出数据清理的分类步骤。

第12章“项目规划”将以前各章的内容合并在一起形成一幅整体性的视图,清晰地表明了如何管理数据仓库项目以及如何指导项目管理者进行项目规划。它还说明了如何在项目执行过程中进行控制的指导方式和劝诫手段。

致谢

每一本书的出版都离不开许多人:家人、朋友、商业伙伴、校对人员以及出版商的支持和奉献。

我们首先要感谢我们的家人,没有他们的关心,这本书不可能与大家见面。我们感谢 Milinda Ann Smith 先生为我们留出清静的工作空间。特别要感谢 Donald Philip Sherman 花费时间检查每一章,修改可读性与语法问题。结果,他本人也成为了数据仓库的准专家。

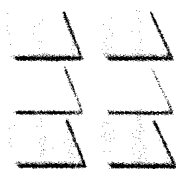
我们也要感谢我们的朋友和商业伙伴,他们在百忙之中抽出时间来审阅稿件的内容和清晰度。我们得到了许多有价值的意见,在此特别感谢提出意见的 Joyce Bischoff, Lee Donoso, David Foote, Pat Higgs, David Marco, Rachel

Meyers, Frank McGuff, Pieter Mimno, Dori Ann Neuman, Clay Rehm, J. D. Welch 和 Duncan Witte 等朋友。

我们特别要感谢 Jane Aubol 和 Ed Carstens, 他们为完成书中的一些章节付出了辛勤的劳动。他们的努力对提高本书的质量产生了巨大影响。Jane 女士不仅花费了大量时间阅读每一章并提出改进意见, 而且还亲自撰写了部分章节。我们同样感谢 Ed 先生对于“项目规划”一章提出的宝贵意见。

我们也要感谢我们的出版商, Addison-Wesley, 是他使我们的梦想成真。我们感谢 Elizabeth Spainhour 和 Mary O'Brien 鼓励我们走出的每一步并为我们搭桥铺路, Mariann Kourafas 帮助我们编辑、协调检查, Sarah Weaver 帮助我们核对每一章的内容, 使之达到出版要求, Jacquelyn Doucette 管理作品内容, Curt Johnson 在许多国家为本书开拓市场。我们感谢 Addison-Wesley 出版社中为此书工作过的所有员工。

最后, 我们还要感谢加州电影城的 Jerry's Famous Deli, 它为我们提供了最好的酒店和丰盛的餐饮。为我们的大量书籍例会提供了宽松舒适的氛围。



目 录

3	策划人语
7	译者序
11	英文版前言
1	第 1 章 数据仓库概述
3	传统开发模式
7	数据仓库开发模式
14	项目管理的任务
17	数据仓库项目管理的难点
20	提要
21	练习
31	第 2 章 项目目标
32	传统决策支持系统(DSS)的固有缺陷
40	数据管理的解决方案
44	短期目标