

# 软件评估、 基准测试与最佳实践

## Software Assessments, Benchmarks, and Best Practices

(美) Capers Jones 著  
韩柯 等译



机械工业出版社  
China Machine Press



中信出版社  
CITIC PUBLISHING HOUSE

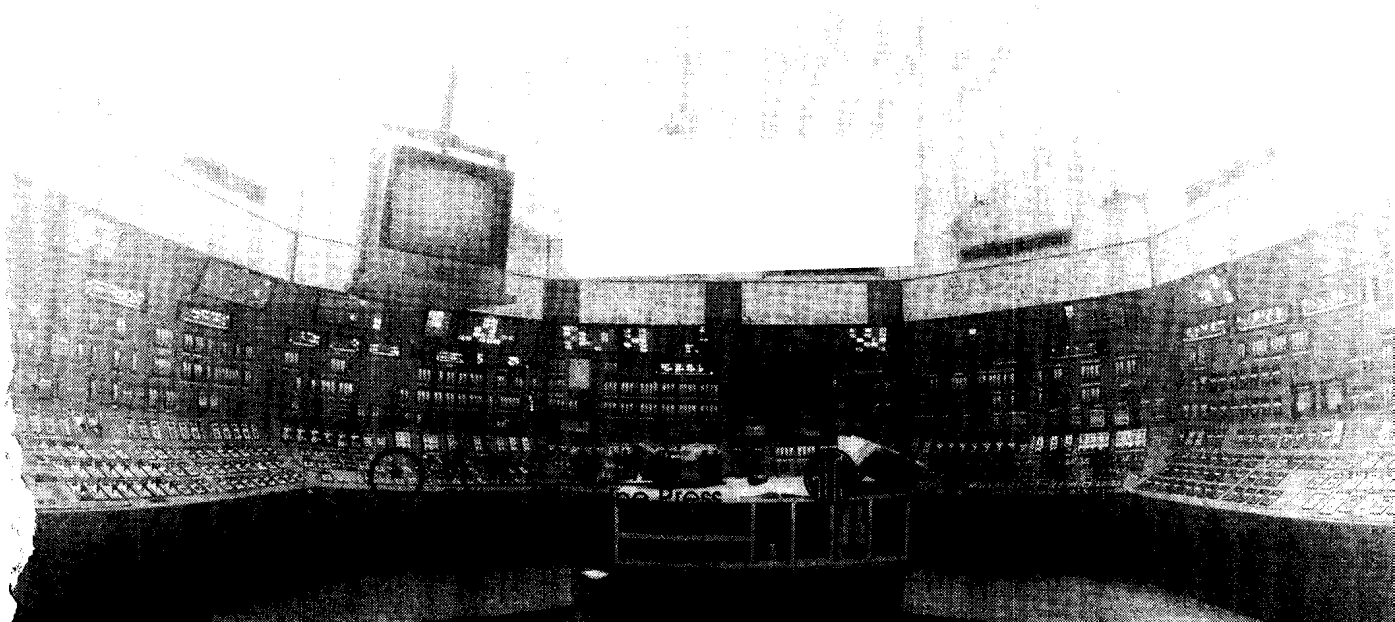
71'311.52  
678

# 软件评估、 基准测试与最佳实践

Software Assessments,  
Benchmarks, and Best Practices

(美) Capers Jones 著

韩柯 等译



本书不仅详细介绍了软件评估和基准测试的方法以及如何有针对性地改进软件评估和基准测试,而且还给出了美国在实际软件工程各个主要方面的大量数据,对于改进我们的软件项目具有很高的参考价值。

本书的突出特点是将定性与定量方法结合起来分析软件开发活动。结合分析评估数据和基准测试数据,可以揭示特定的工具和实践如何影响机构开发工作的效率,研究结果提供了更加清晰和更加宏观的指导,使机构可以确定改进开发工作的领域。

本书不仅可以作为软件企业改善开发工作的指导,而且也可以作为软件学院相关课程的教材或参考书。

Authorized translation from the English language edition entitled *Software Assessments, Benchmarks, and Best Practices* (ISBN: 0-201-48542-7) by Capers Jones, published by Pearson Education, Inc., publishing as Addison Wesley Longman, Copyright © 2000 by Addison Wesley Longman, Inc.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanic, including photocopying, recording, or by any information storage retrieval system, without permission of Pearson Education, Inc.

Chinese simplified language edition published by China Machine Press and CITIC PUBLISHING HOUSE.

Copyright © 2003 by China Machine Press and CITIC PUBLISHING HOUSE.

本书中文简体字版由美国Pearson Education培生教育出版集团授权机械工业出版社与中信出版社合作出版。未经出版者书面许可,不得以任何方式复制或抄袭本书内容。

版权所有,侵权必究。

**本书版权登记号: 图字: 01-2001-5468**

**图书在版编目(CIP)数据**

软件评估、基准测试与最佳实践 / (美)琼斯(Jones, C.)著;韩柯等译 - 北京:机械工业出版社,2003.4

(软件工程技术丛书 质量管理系列)

书名原文: *Software Assessments, Benchmarks, and Best Practices*

ISBN 7-111-11805-7

I. 软… II. ①琼… ②韩… III. 软件开发 - 评估 IV. TP311.52

中国版本图书馆CIP数据核字(2003)第017232号

机械工业出版社(北京市西城区百万庄大街22号 邮政编码 100037)

责任编辑:迟振春

北京牛山世兴印刷厂印刷·新华书店北京发行所发行

2003年4月第1版第1次印刷

787mm × 1092mm 1/16 · 26.5印张

印数: 0 001-5 000册

定价: 49.00元

凡购本书,如有倒页、脱页、缺页,由本社发行部调换

# 译者序

与软件危机一起诞生的软件工程方法和建模理论已经发展了几十年了。但是现实的情况是，软件项目存在的问题仍然很严重，有头无尾的半截子工程、用户不满意的工程、难以投入实际使用的工程、严重超支和拖延进度的工程，并不是个别现象。

从世界范围看，软件生产还属于劳动力高度密集的产业，生产方式是高度专业化的、创造性的个人脑力劳动，具有很浓的工艺色彩，同样的功能往往可以根据开发人员的个人爱好和能力，通过很多种方式实现。软件项目的质量、进度和成本水平基本上取决于软件人员个人的专业素质，以及与转化这些专业素质潜能有关的技术、管理和人员因素。另一方面，同样由于软件具有很浓的工艺色彩，对于具有一定规模的软件，很难在较短的时间内，通过简单的方法全面、准确、合理地测量和比较软件质量。

本书的突出特点是将定性与定量方法结合起来分析软件开发活动。首先把软件项目划分为6个大类，然后综合分析评估数据和基准测试数据，分别确定各类软件的平均水平和最佳水平，科学地揭示特定的工具和实践如何影响机构开发工作的效率，研究结果非常清晰。

本书不仅详细介绍了软件评估和基准测试的方法以及如何有针对性地改进软件评估和基准测试，而且还给出了美国实际软件工程各个主要方面的大量数据，对于改进我们自己的软件项目具有很高的参考价值。

参加本书翻译工作的人员除封面署名外，还有：黄慧菊、屈健、刘芙蓉、王威、李津津、陈秋湖、曹永光、朱军、解冀海、周清、杜蔚轩、原小玲、韩文臣、张玉光、焦跃、孟海军、贺玉寅、黄飞、张捷等。

在翻译过程中，除了对原文个别明显文字错误进行了相应更正外，我们力求忠实原文。但由于译者的知识水平和实际工作经验有限，不当之处在所难免，恳请读者批评指正。

译者

# 前言

在撰写本书的过程中，我最初的写作计划曾经做了很多修改。最初我曾经打算把本书分成两个主要部分，在第一部分讨论美国和欧洲所使用的几种评估和基准测试方法，在第二部分全面论述软件生产率和质量基准测试，以及从基准测试研究中总结出的相关的“最佳实践”。本书的基准测试和最佳实践涵盖六类主要的软件项目：1) 管理信息系统项目；2) 外包项目；3) 系统与嵌入式软件项目；4) 商品化软件项目；5) 军用软件项目；6) 由最终用户开发的个人软件项目。

但是在着手进行写作时，本书的论述焦点发生了变化。我很快就明白过来，完整地讨论这六类软件中的每一类软件的基准测试和最佳实践，所需要的篇幅大约是计划的两倍。我曾经计划用大约30页的篇幅给出每类软件的基准测试和最佳实践信息，但是为了合理地利用已有的数据，六类软件中的五类都需要几乎60页的篇幅。不仅如此，讨论如何展开评估和基准测试研究，以及这些评估和基准测试研究之间的技术差别，对于像我们一样从事评估和基准测试研究的读者来说还是很有用的，但是对于基准测试咨询圈外的读者来说，不一定很有用处。

因此，我对有关评估和基准测试方法的讨论做了一定的压缩，用更多的篇幅讨论在评估和基准测试研究中所收集到的信息。按照本书原来的打算，两部分的篇幅大致相等，最后变成介绍部分大量压缩，并用主要篇幅讨论每类软件以及每类软件所面临的各种问题。

本书还强调了美国的评估和基准测试数据。尽管我和我的同事已经收集了多达24个国家的数据，但是国际基准测试问题相当复杂。工作年和工作日、在如何处理超时工作等问题上不同国家之间的差别，以及欧洲对某些类型数据收集的限制，促使我决心集中讨论美国的数据。

读者应该注意到，本书是由评估和基准测试界的专家所撰写的有关评估和基准测试问题的专著。由于我的公司从1985年以来一直从事评估和基准测试业务，因此对这个主题非常感兴趣。但是，本书不是一本进行市场宣传的书，也不是专门论述我的公司如何开展评估和基准测试工作的。软件评估和基准测试都是非常重要的，本书试图介绍所有评估和基准测试咨询机构所遵循的一般原则。

在我看来，并且我的竞争对手也这样认为，软件评估与软件基准测试对于全球经济是非常重要的。软件已经成为20世纪劳动力最密集的产品，同时也是最容易出错的产品。软件评估、基准测试以及经验数据对于尽可能避免软件项目的失败，具有至关重要的作用。每个软件项目管理人员、软件质量保证专业人员以及软件工程师，都应该理解软件评估和基准测试的基本概念。这是整个评估和基准测试咨询界的共识。

软件界在开发进度预估的精确性、成本计算的精确性以及质量控制上，可谓是声名狼藉。绝大多数大型系统都延期完工、超出预算，很多项目在没有完成之前就被取消了。软件评

估和基准测试，以及在这两项活动基础上进行的有计划的过程改进程序，有助于使软件开发工作处于管理控制之下。这些评估和基准测试并不是解决问题的“撒手铜”，评估、基准测试和过程改进程序都需要投入，而且这种投入还可能很高。但是项目失败的代价要远远超过在这些方面的投入。

本书讨论可以从软件评估和基准测试研究中受益的各种复杂软件项目。小型和简单项目不是评估和基准测试的主要关注点。软件评估、基准测试和过程改进的着眼点以及本书所讨论的重点，都放在大型和复杂应用系统上。

第1章将引入软件评估和基准测试主题，讨论应该收集的数据种类。该章还将介绍需要加以避免的一些常见问题，例如在使用数据之前没有经过核对，以及使用不可靠的度量标准，例如代码行数。该章还将讨论有关保护客户数据的必要性，提出了一些编码方法，在不泄漏客户专用信息的前提下进行基准测试。

第2章将介绍软件过程评估的历史，讨论在软件过程评估期间所收集的某些种类的信息。尽管有十多种形式的软件评估，但是由卡内基·梅隆的SEI（软件工程研究所）所推广的评估形式是最著名的。有些较新或比较专业的评估，例如用于Y2K问题的评估，自从1998年以来也得到广泛应用。

第3章将讨论与软件基准测试和软件基线测试有关的问题。基准测试收集定量数据并与业界标准数据进行比较。基线测试以最初的起点为基准进行测试，测量公司提高生产率和质量的速率。当然，有时生产率和质量还会降低，而不是提高。

第4章将讨论在软件评估和基准测试研究过程中应该记录的36个关键因素。如果记录了这36个关键因素，就可以合理地比较由几乎所有基准测试和评估咨询机构，或由任何公司或政府部门所收集的数据。

第5章将澄清软件界一个模棱两可的重要问题。当我们说“最佳实践”时，确切的含义是什么？第5章将讨论根据最佳实践状态而评判工具和技术的评判标准。作者认为，任何一种被认为是潜在最佳实践的技术，都需要拿出至少在10个公司和50个项目中进行实际应用的经验数据。

第6章将讨论软件评估和基准测试研究的后续活动。软件评估和基准测试两种研究本质上都是诊断式的，而不是治疗式的。这两种研究可以找出问题，但是不能解决问题。因此，进行了评估或基准测试研究之后，接下来的活动自然就应该过程改进程序。

第7章将介绍管理信息系统项目的基准测试和最佳实践。管理信息系统项目是由公司或政府部门为其自己内部使用而构建的软件应用系统。管理信息系统应用系统常常要访问大型公司数据库，其主要用途是将原始数据转换为有用的信息。尽管小型管理信息系统项目的开发生产率一般来说相当高，但是大型管理信息系统项目的失败概率却通常要超过项目平均失败概率。大型管理信息系统项目的质量往往也比较差。

第8章将介绍外包软件项目的基准测试和最佳实践。该章讨论的重点是根据管理信息系统合同展开的项目，而不是军用软件或系统软件外包项目。绝大多数外包项目承包商，例如安达信咨询公司、电子数据系统公司和IBM公司，都把重点放在管理信息系统市场上，因为这是他们外包服务的最大市场。一般来说，与内部管理信息系统项目相比，外包项目

具有更高的生产率和质量水平，但是客户与外包项目承包商之间的纠纷时有发生。

第9章将介绍系统与嵌入式软件项目的基准测试和最佳实践。系统和嵌入式软件是控制物理设备（例如计算机、电话交换机、飞机飞行控制系统或汽车化油器系统）的应用系统。系统软件与物理硬件设备的紧密结合，产生了非常精细的质量控制方法。系统软件界在功能点数超过10 000的大型应用系统方面，具有最好的历史记录。

第10章将介绍商品化软件项目的基准测试和最佳实践。商品化软件应用系统面向大众市场，有些商品化软件应用系统在全球范围内有数以百万计的客户。商品化软件和系统软件领域在操作系统领域有一定的重叠，因为像Windows 98这样的商品化产品，既是系统软件又是商品化软件。商品化软件界需要解决特殊的问题，例如软件包的翻译和本地化、盗版以及针对病毒的广泛的防范。

第11章将讨论军用软件项目的基准测试和最佳实践，重点是美国军方和国防部。军用软件领域在构建大型和复杂应用系统方面做得相当不错，尽管军用软件开发生产率比其他领域的软件开发生产率都低。美国军用标准遗留给国防领域的是一些非常累赘的实践。军用软件领域的规划和规格说明要比规模相当的民用项目大差不多3倍，这主要是因为军事监督需求，而不是项目的技术需求。

第12章将介绍最终用户软件开发的基准测试和最佳实践。进入21世纪，有超过1200万的美国家办公室工作人员知道如何编写计算机程序，如果他们希望自己编写计算机程序，就可以做到。到21世纪中叶，美国能够编写计算机程序的工作人员数量将达到1.25亿。的确，有迹象显示，未来判断文盲的标准不再是看其是否能够读书写字，而是要看其是否具备必要的计算机知识。最终用户应用系统目前还不属于一般的软件评估和基准测试范畴。更重要的是，最终用户应用系统还处于知识产权法保护的范畴之外。随着最终用户应用系统的逐渐增多，为这些不明确的应用系统建立政策和方针是非常重要的。

在编写本书过程中，基于功能点度量标准的基准测试在除军用软件之外的所有软件界，已经处于统治地位，而在军用软件界，基于代码行数的基准测试仍然盛行。本书使用的是功能点度量标准，没有使用代码行数度量标准，以便建立涉及多种程序设计语言的基准测试标准。本书使用由国际功能点用户集团所定义的功能点4.1版作为度量标准。

# 致 谢

与以前一样，我要特别感谢我的妻子Eileen Jones的帮助，是她使我的所有书得以出版。Eileen负责我们所有的出版合同。她现在已经可以理解这些合同的细节，也认识了一些律师。我还要感谢她在我忙于写作甚至放弃节假日时所表现出的耐心。

本书的手稿是在软件生产率研究所（Software Productivity Research, SPR）和Artemis管理系统公司合并的大约两年之后完成的。我要特别感谢Artemis公司的总裁Steve Yager对SPR评估和基准测试研究的支持，感谢世界各地的很多Artemis公司同事和同行。

我还要感谢Gores集团的Alec Gores和Vance Diggins将SPR纳入这个杰出的机构中。

我要特别感谢SPR总裁Charles Douglass的多年领导和友谊，感谢我的SPR同事们在收集数据方面所给予的帮助。我要特别感谢SPR员工的家庭和朋友，他们需要忍耐家人的经常出差。我要感谢Ed Begley、Chuck Berlin、Amy Bowers、Julie LeBaron、Michael Bragen、Doug Brindley、Jack Boyle、Tom Cagley、Sudip Chakraborty、Michael Cunnane、Gail Flaherty、Richard Gazorian、Mike Griffin、David Gustafson、Bill Harmon、Bob Haven、Steve Hone、Jan Huffman、Peter Katsoulas、John Mulcahy、Joy Mohr、Donna O'Donnell、Mark Pinis、Mickie Prault、Tom Riesmeyer、Evelyn Rooney、Kathy Russell、Todd Santora、Keith Savage、Pam Simonvich、Bill Walsh和John Zimmerman。

我要特别感谢功能点（function point）的发明者Allan Albrecht，感谢他对业界的无价贡献以及在SPR的杰出工作。如果没有Allan在功能点上的开创性工作，创建精确的基线测试和基准测试的能力可能就不存在了。感谢国际功能点用户集团（International Function Point Users Group, IFPUG）对功能点角色的扩展。

我还要感谢Mary O'Brien及其Addison Wesley Longman出版公司的同事。能够与杰出的编辑和出版团队一起工作，对于每一位作者来说永远都是一种享受。

我要感谢对软件评估、基准测试与基线测试、测量以及过程改进感兴趣的客户机构，这种兴趣使得我们能够在一起工作。正是这些机构的数据使得预估工具得以产生。

要列举的名字实在太多，不过我要特别感谢安达信咨询公司、AT&T公司、贝尔亚特兰大公司、贝尔北方研究所、贝尔Sygma公司、Bendix公司、英国航空公司、CBIS公司、Charles Schwab公司、近代圣徒教堂、辛辛那提贝尔公司、CODEX公司、瑞士信用、DEC公司、Dun & Bradstreet公司、杜邦公司、电子数据系统（EDS）公司、Finsiel公司、福特汽车公司、Fortis集团、通用电气公司、通用汽车公司、GTE公司、哈特福德保险公司、惠普公司、IBM公司、Informix公司、Inland钢铁公司、美国国家税务局、JC Penney公司、JP Morgan公司、Kozo Keikaku公司、语言技术公司、Litton公司、Lotus公司、朗讯公司、Mead数据中心、麦肯锡咨询公司、微软公司、摩托罗拉公司、日本电报公司、NCR公司、北方电信公司、大西洋贝尔公司、太平洋贝尔公司、Ralston Purina公司、Sapiens公司、

## VIII

Sears Roebuck 公司、西门子利多富公司、软件出版集团公司、SOGEI 公司、太阳人寿保险公司、Tandem 公司、TRW 公司、Unisys 公司、美国空军、美国海军水面武器集团、美国西部公司、西屋公司等很多其他公司的同事和客户。

我还要感谢软件评估和基准测试研究方面的同事和竞争对手。如果没有像Howard Rubin、Watts Humphrey、Bill Curtis、Ron Radice、Mark Paulk 以及Bob Park 这些研究人员的前期工作，评估和基准测试领域就不会受到软件界的普遍关注。我要特别感谢IBM 公司已故的Ted Climis 和Jim Frame，他们是将软件评估和基准测试引入IBM 公司软件部门的关键人物。

## 摘要

**基准数据**是用于将企业与同行业其他类似企业进行比较的定量数据集合。软件基准数据常用于生产率和质量比较，也用于工资和成本水平的比较。

**基线数据**是用于标记过程改进程序起点的定量数据集合。基线数据和基准数据在收集定量数据方面很相似。基线测试常常用于外包协议，用于确定提高生产率和质量的合同义务的起点。

**软件评估**是用于开发软件应用系统的方法、工具和过程的现场评审。评估数据在本质上是定性的。评估用于将机构置于可比较的位置上，并给出其能力的相对水平。

如果将评估数据、基准测试数据和基线测试数据结合起来，则会产生很强的综合结果。评估、基准测试和基线测试的结合可以用来找出最佳实践、平均实践和最差实践。术语最佳实践是指在至少**10**个公司和**50**个项目的重复试验中，将软件质量和生产率水平在平均水平的基础上使方法论、工具软件包和组织结构至少提高**15 %**。

# 目 录

译者序

前言

致谢

摘要

## 第1章 概述 .....1

- 1.1 保护和比较机密数据 .....2
- 1.2 国家与城市代码 .....5
- 1.3 使用标准行业分类代码进行软件研究 .....5
- 1.4 用于基准测试和评估研究的软件项目  
分类 .....7

## 第2章 软件过程评估 .....13

- 2.1 软件过程评估的起源 .....15
- 2.2 SEI评估方法 .....18
- 2.3 SPR评估方法 .....21
  - 2.3.1 客户的优势模式（优于平均水平） .....25
  - 2.3.2 平均水平模式（具有行业平均水平） .....25
  - 2.3.3 客户的弱势模式（低于平均水平） .....26
- 2.4 SEI与SPR等级评定系统的相关性 .....28
- 2.5 有关软件评估的补充读物和参考资料 .....29

## 第3章 软件基准测试与基线测试 .....35

- 3.1 基准测试与尺寸度量标准的问题 .....45
- 3.2 基准测试软件成本 .....56
- 3.3 使用项目级数据进行软件基准测试  
研究的风险 .....62
- 3.4 软件基准测试研究中邮件调查的风险 .....63
- 3.5 向基于活动的软件基准测试数据发展 .....64
- 3.6 软件质量基准测试 .....66
- 3.7 软件基线测试 .....69
- 3.8 有关软件基准测试与基线测试的  
补充读物和参考资料 .....70

## 第4章 软件评估与基准测试研究的36个

关键因素 .....77

- 4.1 软件分类因素 .....77
- 4.2 与特定项目有关的因素 .....78
- 4.3 技术因素 .....78
- 4.4 社会因素 .....79
- 4.5 人体工程学因素 .....79
- 4.6 国际因素 .....80
- 4.7 有关因素分析的补充读物和参考资料 .....81

## 第5章 标识软件最佳与最差实践 .....87

- 5.1 影响软件开发生产率的因素 .....90
- 5.2 影响软件维护生产率的因素 .....92
- 5.3 积极与消极因素的模式 .....94
- 5.4 有关软件最佳与最差实践的补充读物和  
参考资料 .....96

## 第6章 软件过程改进 .....97

- 6.1 年度战术与战略软件改进计划 .....98
- 6.2 年度软件进展报告 .....98
- 6.3 通往优秀软件之路的6个阶段 .....99
  - 6.3.1 第0阶段：软件过程评估、基线  
测试与基准测试 .....99
  - 6.3.2 第1阶段：注重管理技术 .....101
  - 6.3.3 第2阶段：注重软件过程和方法 .....101
  - 6.3.4 第3阶段：注重新的工具和途径 .....102
  - 6.3.5 第4阶段：注重基础设施和专门化 .....103
  - 6.3.6 第5阶段：注重可复用性 .....104
  - 6.3.7 第6阶段：注重行业领先 .....104
- 6.4 过程改进的成本、时间安排与价值 .....105
- 6.5 有关软件过程改进的补充读物和参考  
资料 .....107

## 第7章 管理信息系统软件项目的基准 测试与最佳实践 .....111

|                                           |     |                                         |     |
|-------------------------------------------|-----|-----------------------------------------|-----|
| 7.1 管理信息系统与公司数据库 .....                    | 112 | 7.11.3 管理信息系统软件的最佳管理人员<br>培训与教育实践 ..... | 143 |
| 7.2 缺少数据度量标准 .....                        | 112 | 7.11.4 管理信息系统软件领域的最佳<br>专门化实践 .....     | 144 |
| 7.3 管理信息系统与企业资源规划 .....                   | 113 | 7.11.5 管理信息系统软件的最佳报酬<br>与工资实践 .....     | 144 |
| 7.4 管理信息系统、欧元与2000年问题 .....               | 114 | 7.11.6 管理信息系统软件的最佳办公室<br>人体工程学实践 .....  | 145 |
| 7.5 管理信息系统技术 .....                        | 114 | 7.11.7 管理信息系统软件的最佳机构结构<br>实践 .....      | 146 |
| 7.6 信息系统与系统软件的不同起源 .....                  | 117 | 7.11.8 管理信息系统软件的最佳员工士气<br>实践 .....      | 147 |
| 7.7 管理信息系统的统计数据 .....                     | 119 | 7.11.9 管理信息系统软件的最佳工作模式与<br>超时工作实践 ..... | 148 |
| 7.8 管理信息系统基准测试 .....                      | 122 | 7.11.10 减少管理信息系统员工跳槽率的<br>最佳实践 .....    | 149 |
| 7.9 管理信息系统软件的成功与失败 .....                  | 127 | 7.12 管理信息系统软件小结与结论 .....                | 149 |
| 7.9.1 管理信息系统的成功因素 .....                   | 129 | 7.13 有关管理信息系统软件的补充读物和<br>参考资料 .....     | 150 |
| 7.9.2 管理信息系统的失败因素 .....                   | 130 | <b>第8章 外包软件的基准测试与最佳实践</b> .....         | 155 |
| 7.10 管理信息系统软件的最佳技术实践 .....                | 130 | 8.1 国际外包的增长 .....                       | 157 |
| 7.10.1 管理信息系统软件的最佳项目<br>管理实践 .....        | 131 | 8.2 选择外包或内部开发 .....                     | 159 |
| 7.10.2 管理信息系统软件的最佳需求收集<br>与分析实践 .....     | 133 | 8.3 降低外包承包商及其客户之间产生<br>矛盾的风险 .....      | 164 |
| 7.10.3 管理信息系统软件的最佳设计与<br>规格说明实践 .....     | 134 | 8.4 外包技术 .....                          | 167 |
| 7.10.4 管理信息系统软件的最佳编码<br>实践 .....          | 135 | 8.5 外包统计数据 .....                        | 170 |
| 7.10.5 管理信息系统软件的最佳复用<br>实践 .....          | 137 | 8.6 外包基准测试 .....                        | 173 |
| 7.10.6 管理信息系统软件的最佳变更控制<br>实践 .....        | 138 | 8.7 外包的成功与失败 .....                      | 178 |
| 7.10.7 管理信息系统软件的最佳用户文档<br>实践 .....        | 138 | 8.7.1 外包与合同软件的成功因素 .....                | 181 |
| 7.10.8 管理信息系统软件的最佳质量控制<br>与缺陷预先清除实践 ..... | 139 | 8.7.2 外包与合同软件的失败因素 .....                | 181 |
| 7.10.9 管理信息系统软件的最佳测试实践<br>与工具 .....       | 140 | 8.8 外包软件的最佳技术实践 .....                   | 182 |
| 7.10.10 管理信息系统软件的最佳维护与<br>增强实践 .....      | 141 | 8.8.1 外包软件的最佳项目管理实践 .....               | 182 |
| 7.11 管理信息系统软件的最佳人员实践 .....                | 141 | 8.8.2 外包软件的最佳需求收集与分析<br>实践 .....        | 185 |
| 7.11.1 管理信息系统软件的最佳员工雇用<br>实践 .....        | 142 | 8.8.3 外包软件的最佳设计与规格说明<br>实践 .....        | 186 |
| 7.11.2 管理信息系统软件的最佳员工培训<br>与教育实践 .....     | 142 | 8.8.4 外包软件的最佳编码实践 .....                 | 187 |

|                                      |     |                                        |     |
|--------------------------------------|-----|----------------------------------------|-----|
| 8.8.5 外包软件的最佳复用实践 .....              | 188 | 9.4.5 系统软件的最佳复用实践 .....                | 230 |
| 8.8.6 外包软件的最佳变更控制实践 .....            | 190 | 9.4.6 系统软件的最佳变更控制实践 .....              | 232 |
| 8.8.7 外包软件的最佳用户文档实践 .....            | 191 | 9.4.7 系统软件的最佳用户文档实践 .....              | 233 |
| 8.8.8 外包软件的最佳质量控制与缺陷预先<br>清除实践 ..... | 192 | 9.4.8 系统软件的最佳质量控制与缺陷预先<br>清除实践 .....   | 234 |
| 8.8.9 外包软件的最佳测试实践与工具 .....           | 193 | 9.4.9 系统软件的最佳测试实践与工具 .....             | 237 |
| 8.8.10 外包软件的最佳维护与增强实践 .....          | 194 | 9.4.10 系统软件的最佳维护与增强实践 .....            | 237 |
| 8.9 外包软件的最佳人员实践 .....                | 195 | 9.5 系统软件的最佳人员实践 .....                  | 238 |
| 8.9.1 外包软件的最佳员工雇用实践 .....            | 195 | 9.5.1 系统软件的最佳员工雇用实践 .....              | 239 |
| 8.9.2 外包软件的最佳员工培训与教育<br>实践 .....     | 196 | 9.5.2 系统软件的最佳员工培训与教育<br>实践 .....       | 240 |
| 8.9.3 外包软件的最佳管理人员培训<br>与教育实践 .....   | 197 | 9.5.3 系统软件的最佳管理人员培训与教育<br>实践 .....     | 241 |
| 8.9.4 外包软件领域的最佳专门化实践 .....           | 197 | 9.5.4 系统软件领域的最佳专门化实践 .....             | 242 |
| 8.9.5 外包软件的最佳报酬与工资实践 .....           | 198 | 9.5.5 系统软件的最佳报酬与工资实践 .....             | 242 |
| 8.9.6 外包软件的最佳办公室人体工程学<br>实践 .....    | 199 | 9.5.6 系统软件的最佳办公室人体工程学<br>实践 .....      | 242 |
| 8.9.7 外包软件的最佳机构结构实践 .....            | 199 | 9.5.7 系统软件的最佳机构结构实践 .....              | 244 |
| 8.9.8 外包软件的最佳员工士气实践 .....            | 200 | 9.5.8 系统软件的最佳员工士气实践 .....              | 245 |
| 8.9.9 外包软件的最佳工作模式与超时<br>工作实践 .....   | 201 | 9.5.9 系统软件的最佳工作模式与超时工作<br>实践 .....     | 247 |
| 8.9.10 减少外包员工跳槽率的最佳实践 .....          | 202 | 9.5.10 减少系统软件员工跳槽率的最佳<br>实践 .....      | 247 |
| 8.10 外包软件小结与结论 .....                 | 202 | 9.6 系统软件小结与结论 .....                    | 248 |
| 8.11 有关外包软件的补充读物和参考资料 .....          | 203 | 9.7 有关系统软件的补充读物和参考资料 .....             | 248 |
| <b>第9章 系统软件的基准测试与最佳实践</b> .....      | 207 | <b>第10章 商品化软件的基准测试与最佳<br/>实践</b> ..... | 255 |
| 9.1 系统软件统计数据 .....                   | 208 | 10.1 商品化软件的特殊问题与技术 .....               | 256 |
| 9.2 系统软件基准测试 .....                   | 212 | 10.2 商品化软件统计数据 .....                   | 260 |
| 9.3 系统软件的成功与失败 .....                 | 219 | 10.3 商品化软件基准测试 .....                   | 263 |
| 9.3.1 系统软件的成功因素 .....                | 220 | 10.4 商品化软件的成功与失败 .....                 | 269 |
| 9.3.2 系统软件的失败因素 .....                | 221 | 10.4.1 商品化软件的成功因素 .....                | 271 |
| 9.4 系统软件的最佳技术实践 .....                | 223 | 10.4.2 商品化软件的失败因素 .....                | 272 |
| 9.4.1 系统软件的最佳项目管理实践 .....            | 223 | 10.5 微软公司模式 .....                      | 273 |
| 9.4.2 系统软件的最佳需求收集与<br>分析实践 .....     | 226 | 10.6 ISO 9000—9004标准的出现 .....          | 274 |
| 9.4.3 系统软件的最佳设计与规格<br>说明实践 .....     | 227 | 10.7 商品化软件的最佳技术实践 .....                | 274 |
| 9.4.4 系统软件的最佳编码实践 .....              | 228 | 10.7.1 商品化软件的最佳项目管理实践 .....            | 275 |

|                                        |            |                                       |     |
|----------------------------------------|------------|---------------------------------------|-----|
| 10.7.2 商品化软件的最佳需求收集与分析<br>实践 .....     | 277        | 11.3 Airlie Council .....             | 306 |
| 10.7.3 商品化软件的最佳设计与规格说明<br>实践 .....     | 278        | 11.4 军用软件统计数据 .....                   | 307 |
| 10.7.4 商品化软件的最佳编码实践 .....              | 278        | 11.5 军用软件基准测试 .....                   | 310 |
| 10.7.5 商品化软件的最佳复用实践 .....              | 279        | 11.6 军用软件的成功与失败 .....                 | 317 |
| 10.7.6 商品化软件的最佳变更控制实践 .....            | 281        | 11.6.1 军用软件的成功因素 .....                | 317 |
| 10.7.7 商品化软件的最佳用户文档实践 .....            | 281        | 11.6.2 军用软件的失败因素 .....                | 318 |
| 10.7.8 商品化软件的最佳质量控制与缺陷<br>预先清除实践 ..... | 283        | 11.7 军用软件的最佳技术实践 .....                | 320 |
| 10.7.9 商品化软件的最佳测试实践<br>与工具 .....       | 284        | 11.7.1 军用软件的最佳项目管理实践 .....            | 320 |
| 10.7.10 商品化软件的最佳维护与增强<br>实践 .....      | 285        | 11.7.2 军用软件的最佳需求收集与分析<br>实践 .....     | 323 |
| 10.8 商品化软件的最佳人员实践 .....                | 287        | 11.7.3 军用软件的最佳设计与规格说明<br>实践 .....     | 325 |
| 10.8.1 商品化软件的最佳员工雇用实践 .....            | 287        | 11.7.4 军用软件的最佳编码实践 .....              | 326 |
| 10.8.2 商品化软件的最佳员工培训与教育<br>实践 .....     | 288        | 11.7.5 军用软件的最佳复用实践 .....              | 327 |
| 10.8.3 商品化软件的最佳管理人员培训与<br>教育实践 .....   | 289        | 11.7.6 军用软件的最佳变更控制实践 .....            | 329 |
| 10.8.4 商品化软件领域的最佳专门化实践 .....           | 289        | 11.7.7 军用软件的最佳用户文档实践 .....            | 330 |
| 10.8.5 商品化软件的最佳报酬与工资实践 .....           | 290        | 11.7.8 军用软件的最佳质量控制与缺陷<br>预先清除实践 ..... | 331 |
| 10.8.6 商品化软件的最佳办公室人体工程学<br>实践 .....    | 290        | 11.7.9 军用软件的最佳测试实践与工具 .....           | 332 |
| 10.8.7 商品化软件的最佳机构结构实践 .....            | 291        | 11.7.10 军用软件的最佳维护与增强实践 .....          | 333 |
| 10.8.8 商品化软件的最佳员工士气实践 .....            | 292        | 11.8 军用软件的最佳人员实践 .....                | 334 |
| 10.8.9 商品化软件的最佳工作模式与超时<br>工作实践 .....   | 293        | 11.8.1 军用软件的最佳员工雇用实践 .....            | 335 |
| 10.8.10 减少商品化软件员工跳槽率的最佳<br>实践 .....    | 293        | 11.8.2 军用软件的最佳员工培训与教育<br>实践 .....     | 335 |
| 10.9 商品化软件小结与结论 .....                  | 294        | 11.8.3 军用软件的最佳管理人员培训<br>与教育实践 .....   | 336 |
| 10.10 有关商品化软件的补充读物和<br>参考资料 .....      | 294        | 11.8.4 军用软件领域的最佳专门化实践 .....           | 337 |
| <b>第11章 军用软件的基准测试与最佳实践</b> .....       | <b>301</b> | 11.8.5 军用软件的最佳报酬与工资实践 .....           | 337 |
| 11.1 军用软件技术与特殊问题 .....                 | 302        | 11.8.6 军用软件的最佳办公室人体<br>工程学实践 .....    | 338 |
| 11.2 软件工程研究所 .....                     | 304        | 11.8.7 军用软件的最佳机构结构实践 .....            | 338 |
|                                        |            | 11.8.8 军用软件的最佳员工士气实践 .....            | 340 |
|                                        |            | 11.8.9 军用软件的最佳工作模式与超时<br>工作实践 .....   | 340 |
|                                        |            | 11.8.10 减少军用软件员工跳槽率的<br>最佳实践 .....    | 341 |
|                                        |            | 11.9 军用软件小结与结论 .....                  | 342 |

11.10 有关军用软件的补充读物和  
参考资料 .....343

**第12章 最终用户软件的基准测试与  
最佳实践 .....351**

12.1 最终用户软件统计数据 .....353  
12.2 最终用户软件基准测试 .....353  
12.3 最终用户软件的成功与失败 .....356  
    12.3.1 最终用户软件的成功因素 .....357  
    12.3.2 最终用户软件的失败因素 .....357

12.4 最终用户软件的最佳技术实践 .....357  
12.5 最终用户软件的最佳人员实践 .....358  
12.6 有关最终用户软件的补充读物和参考  
资料 .....359

**附录 SPR 评估、基准测试与基线  
测试问卷 .....361**

**术语表 .....385**

**参考文献 .....395**

# 第 1 章

## 概 述

---

50多年来,软件一直是一种麻烦的技术。计算机和软件是企业、政府以及军事机构的主要工具,但是计算机和软件却是棘手的、昂贵的并且是很容易出错的。

开发大型软件应用系统尤其棘手。与其他行业的工作相比,软件应用系统开发是劳动密集程度更高的工作。软件项目被取消或延期完成,或超过预算的情况,要比其他现代行业的项目更经常地出现。软件应用系统还非常容易出现错误,其中有些错误会带来巨大的业务损失。例如,2000年(Y2K)软件问题已经成为历史上造成经济损失最大的业务问题,也是影响最广泛的业务问题。

软件的这种棘手特性以及与大型软件应用系统有关的很多失败和延期案例,一直是全世界顶级执行经理的一个心病。因为软件对于现代世界是如此重要,因此寻找降低失败概率并提高成功概率的途径,对于所有大型企业来说都是很有价值的。

减少软件问题的最有效方法,就是研究构建和维护应用系统的所有可能的方法。通过仔细研究定量和定性数据,就能够将有益的实践从有害的实践中分离出来。当前用于探索软件构建的3种主要分析形式是:

- 1) 软件过程评估,检查开发实践。
- 2) 对比业界经验数据的软件基准测试,确定自己的表现。
- 3) 对比软件基线数据,测量自己所取得的进步。

术语过程评估是指检查软件项目构建和维护方式的正式和结构化的方法。评估可以由外部咨询机构执行,也可以由内部人员执行。但是,由于内部人员很难保证客观评估,因此更常见的是由外部咨询机构执行评估。

术语基准测试是指特定公司之间定量数据的比较,也指公司与业界标准数据的更一般的比较。基准测试研究常常由已经从数以百计的公司中收集了大量数据的咨询机构执行。尽管基准测试可以对不同公司进行,但是如果这些公司是直接竞争者,则要想共享他们的专用信息一般是不可行的。因此,更可行的方法是在多个公司之间进行“盲”研究,并对具体数据点的来源保密。

术语基线测试是指收集定量数据,指示过程改进程序中的起点。改进的速度要与基线的起始值进行比较。如果基线数据用于描述合同内容,例如外包协议,则需要非常精确地执行基线测试。由于进入外包协议的公司很少可以在签订协议的开始就能够提供精确的生产率和质量数据,因此基线测试有时会导致以后的纠纷,甚至是法律纠纷。

对于任何准备在软件过程改进程序上进行投入的公司,都应该谨慎地建立初始的定量基线数据。由于过程改进程序相当昂贵,并且可能需要运行3到5年,因此需要使用生产率、进度计划、质量和质量数据的精确基线数据,确定投资回报率(Return On Investment, ROI),并进而说明过程改进程序的价值。如果没有定量的基线数据,软件过程改进程序就有可能失败,因为

根据公司的评判标准，重大投资必须在3到5年的时间内取得明显的正回报率。

软件评估、基线测试和基准测试一般在所收集的数据上会出现一些重叠。过程评估要收集大量有关活动是否被很好执行的定性数据。基准测试和基线测试也收集一些定性数据，但重点是定量数据。项目规模、生产率水平、质量水平、开发进度、工资水平以及员工水平都是一般的定量因素。

尽管软件评估、基准测试和基线测试这三种活动常常单独执行，但是结合评估、基准测试和基线测试三方面的数据可以产生最有效的结果。软件过程评估收集有用的定性数据，但是如果没有定量数据，则很难评价在过程改进程序中所进行的投资是否取得回报。基准测试和基线测试收集定量数据，但是如果没有来自评估的定性数据，就不能理解为什么生产率和质量水平低于期望值。结合评估、基准测试和基线测试三方面的数据，就能够回答公司最高管理人员和软件执行经理的以下六个重要问题：

- 1) 我们的生产率和质量是优于还是劣于我们的竞争对手？
- 2) 如果我们在某些关键领域落后于竞争对手，那么我们可以做些什么才能改进这种状况？
- 3) 如果我们要在改进上投入，那么需要多大的投入？
- 4) 如果我们要在改进上投入，那么需要在多少年内投入？
- 5) 如果我们要在改进上投入，那么投资回报率如何？
- 6) 执行改进程序之后，我们的生产率和质量会达到怎样的水平？

尽管软件评估、基准测试和基线测试单独进行都是很有用的，但是评估、基准测试和基线测试之间有着很强的综合作用。只进行软件评估或只进行软件基准测试研究的公司可以获得有用数据，但只是部分数据。进行评估和基准测试的公司则能够很好地支持软件过程改进程序，并确立最终目标。但是，这种目标是针对最初定量基线数据的相对目标。

在编写本书时，《财富》杂志500强中的大多数大型公司都已经完成了其主要软件开发团队的软件过程评估工作。少数公司已经完成了软件基准测试和基线测试研究调查，并且这个百分比正在提高。我的公司（软件生产率研究所[SPR]）已经进行了《财富》杂志500强中大约150家公司的软件评估和基准测试研究调查。我们还为100多家较小的公司，以及相当数量的政府部门和军队机构进行了软件评估和基准测试研究调查。由于很多大型企业都有多个软件开发地点和实验室，所以我们从差不多600个场所收集数据。

由于软件评估、基准测试和基线测试常常由不同的公司使用不同的方法执行，因此在不同的研究结果之间进行比较并不容易。本书推荐的一种方法是软件评估、基准测试和基线测试中的关键数据要素开发一种国际标准，这种标准使得软件评估、基准测试和基线测试数据能够进行比较，并且最终能够导致很多不同来源的数据融合。

## 1.1 保护和比较机密数据

在从大约600个场所对大约9000个软件项目进行数据收集和分析过程中，我们必须处理与大规模研究有关的实际问题。例如，所收集的一部分数据是专用和保密的。我们还曾遇到有的客户同时使用两家或更多评估和基准测试咨询机构的情况。客户自然对两家咨询公司所收集的数