

本书仅提供部分阅读，如需完整版，请联系QQ: 2011705918

提供各种IT类, 其它类书籍pdf下载，如有需要，请QQ:2011705918

注：链接至淘宝，整理那么多资料也不容易，请多多见谅！非诚勿扰！

[点击购买完整版](#)

# DBA修炼之道

## 数据库管理员的第一本书

(原书第2版)

[美] Craig S. Mullins 著 褚晓霞 译

---

### Database Administration

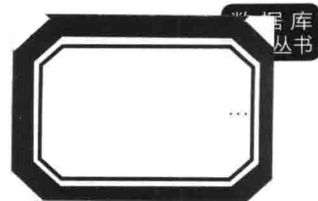
The Complete Guide to DBA Practices and Procedures, Second Edition

---

- 资深数据库管理专家30余年经验结晶，内容全面而深入，用清晰易懂的语言阐释数据库管理的各种技术细节、方法和最佳实践
- 包含大量典型数据管理分析案例，为快速掌握并灵活运用数据库管理技术提供最佳指导



机械工业出版社  
China Machine Press



# DBA修炼之道

## 数据库管理员的第一本书

(原书第2版)

---

Database Administration

The Complete Guide to DBA Practices and Procedures, Second Edition

---

[美] Craig S. Mullins 著 褚晓霞 译



机械工业出版社  
China Machine Press

## 图书在版编目 (CIP) 数据

DBA 修炼之道: 数据库管理员的第一本书 (原书第 2 版) / (美) 穆林斯 (Mullins C. S.) 著; 褚晓霞译. —北京: 机械工业出版社, 2014.8

(数据库技术丛书)

书名原文: Database Administration: The Complete Guide to DBA Practices and Procedures, Second Edition

ISBN 978-7-111-47395-4

I. D… II. ①穆… ②褚… III. 关系数据库 IV. TP311.138

中国版本图书馆 CIP 数据核字 (2014) 第 161669 号

本书版权登记号: 图字: 01-2013-1803

Authorized translation from the English language edition, entitled *Database Administration: The Complete Guide to DBA Practices and Procedures*, 2E, 978-0-321-82294-9 by Craig S. Mullins, published by Pearson Education, Inc., Copyright © 2013.

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

Chinese simplified language edition published by Pearson Education Asia Ltd., and China Machine Press Copyright © 2014.

本书中文简体字版由 Pearson Education (培生教育出版集团) 授权机械工业出版社在中华人民共和国境内 (不包括中国台湾地区和中国香港、澳门特别行政区) 独家出版发行。未经出版者书面许可, 不得以任何方式抄袭、复制或节录本书中的任何部分。

本书封底贴有 Pearson Education (培生教育出版集团) 激光防伪标签, 无标签者不得销售。

## DBA 修炼之道: 数据库管理员的第一本书 (原书第 2 版)

出版发行: 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: 秦 健

责任校对: 董纪丽

印 刷: 蕺城市京瑞印刷有限公司

版 次: 2014 年 9 月第 1 版第 1 次印刷

开 本: 186mm × 240mm 1/16

印 张: 37

书 号: ISBN 978-7-111-47395-4

定 价: 129.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzsj@hzbook.com

版权所有 • 侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

## *Praise* 对本书的赞誉

我已经不记得多少次向人推荐这本书了。这本书写得好，重要的是，它涵盖了成为一名高效 DBA 所需要知道的所有方面。

——Scott Ambler，敏捷数据方法的领导者

这是一本精心编写和组织的数据库管理实践的指南。不像一般的数据库理论或关系数据库理论书籍，这本书更直接地侧重于当今数据库专业人员进行数据库管理的理论与实践，并且没有太多迎合任何特定的产品实施。因此，数据库管理非常适合那些有兴趣研究 DBA 或类似但更具体的角色（如数据建模或数据库性能分析师）的职责的人。

——Sal Ricciardi，微软项目经理

Craig 的特点之一是他的写作清晰、易读。任何技术类书籍的主要目的是在作者和读者之间传递信息，这一点 Craig 做得很棒。他希望读者学习，并且完整表达了自己的想法。

——Chris Foot，远程 DBA 专家和 Oracle ACE

本书涉及 DBA 任务和职责的完整且全面的清单，从创建数据库环境到数据仓库管理，以及中间可能会发生的一切。

——Mike Tarrani，计算机顾问

我认为，每一名业务经理和 IT 经理都应该拥有这样一本书。

——Dan Hotka，独立顾问和 Oracle ACE

Craig Mullins 的这本书见解深刻。Mullins 以卓越的洞察力清晰描述了现代企业中数据管理员和数据库管理员的角色以及他们的职责。

——Michael Tozer，作家及前美国海军军官

## 译者序 *The Translator's Words*

应该说，数据库管理不是什么新概念，随着它的不断发展，数据库管理员（DBA）的需求也在不断增加。DBA 这一职位揭开了数据库管理略带神秘的面纱，为越来越多的公司和人所熟知。

关于数据库管理的书籍不少，其中也不乏优秀和经典之作。在众多数据库管理书籍中，本书仍具有不可替代的地位，我认为主要原因如下。

- 作者。本书的作者 Craig S. Mullins 是数据库管理领域的绝对资深人士，在该领域有超过 30 年的工作经验，涉及数据库的研发和教育、系统分析和设计、数据库和系统管理、数据分析和建模等，本书是他的两大畅销书之一。
- 内容编排。各章按照数据库管理员实际执行工作的顺序进行编排，更加便于想成为 DBA 的人学习，或已经是 DBA 的人查阅参考。
- 内容描述。有别于一般的数据库理论或关系数据库理论的书籍，本书把描述的重点放在当今数据库管理人员实际会用到的理论和实践。
- 与数据库产品无关。本书的独特之处是不以市面上任何一种数据库产品为例来介绍数据库管理的技术和实践。所以具有更加广泛的适用人群。
- 事无巨细。首先，本书详细讲述了数据库管理工作的方方面面，并且在每章的末尾对该章的内容进行小结，同时提供了一些题目供读者思考和实践。书中好多地方都体现出作者的贴心之处，比如在书的开始就以 DBA 和 IT 相关职位的薪资对比告诉读者 DBA 工作的特殊性，以及会遇到的挑战和受到的压力。稍后又专门提供了常用的 DBA 工具，以及 DBA 的经验法则。同时提供了常见的 DBMS 和 DBA 工具供应商，以及一些 DBA 网络资源。更让人觉得贴心的是，作者还给出了一篇 DBA 招聘的示例，供相关企业招聘 DBA 或有意应聘 DBA 的人员参考使用。

从本书的名字就能了解到，这是一本初级读本，但是它绝对不是一本轻松的入门教材。一是因为 DBA 的工作任务烦琐且需要和不同的人打交道，二是因为 DBA 的工作实践比理论更重要。所以，一个 DBA 初学者也许能够很快读懂它，但要想真正掌握还需要长时间的实

践，而本书可以在实践过程中助他们一臂之力。总之，这是一本很好的学习和参考书籍，值得你反复阅读。

翻译这本书，完全出于机缘巧合。应该说，这是我真正意义上翻译的第一本书，翻译这样一本大部头的确是一件很辛苦的事情，所以首先我要感谢机械工业出版社的编辑杨福川，让我有机会把这么优秀的书翻译出来，献给对数据库管理感兴趣的各位朋友。感谢本书的编辑给我提出了不少非常好的意见和建议，使我受益匪浅，感谢对我延迟交稿的容忍。还要感谢我的家人对我的理解和支持，包括精神上的和实际行动上的，我发现这是不做或少做家务的一个很好的理由。最后感谢朋友周移民的帮助，使我最终完成了这本书的翻译工作。

如果这本书能够帮到你，这是我最大的荣幸！

## 前 言 *Preface*

相比 2002 年写这本书的第 1 版时，现在对数据库管理的需求大有增强之势。关系数据库管理系统仍处于生产系统的核心，它们仍然需要数据库管理员的管理。无论你使用的是 Oracle、Microsoft SQL Server、DB2、Informix、Sybase、MySQL、Teradata、PostgreSQL、Ingres，或者是这些流行的 DBMS 产品的任意组合，你都会从本书中受益匪浅。

但是在信息技术领域，十年就是永远。即使一些基本的东西保持不变（如数据库需要管理），很多事情都在变化。这本书的第 2 版收录了在过去的十年中这个行业所发生的对数据库管理产生影响的许多变化。这也是本书的独特之处。同时，它还是业界唯一不基于任何产品来介绍数据库管理技术和实践的书籍。本书定义了数据库管理员的工作，并且用清晰、简单易懂的语言概述了哪些是数据库管理员（或 DBA）所必需的。本书可以：

- 用作学习数据库管理准则的教材。
- 用作建立 DBA 组的依据。
- 用来增加特定 DBMS 的手册或教材。
- 用来帮助高层管理人员理解什么是 DBA，以及为什么需要这个职位。

新加了哪些内容？这个版本的重大改进之一是覆盖合规性。政府和行业法规的数量在过去的十年呈爆炸式增长，并且许多法规规定了数据管理、处理和加工方式的变化。其中最明显的政府法规无疑是 Sarbanes-Oxley Act（又名《美国公众公司会计改革和投资者法案》，2002 年），还有许多其他的法案，包括 HIPAA 和 GLB（Gramm-Leach-Bliley 法案）。最明显的行业法规是 PCI DSS（支付卡行业数据安全标准）。所有这些法规和很多其他法规一起，给数据施加了额外的行政负担。本书用了一整章专门讨论这个话题，包括监管合规性对数据管理任务的影响，例如，元数据管理、数据质量、数据库审计、数据屏蔽（data mask）<sup>⊖</sup>、数据库归档，以及更传统的 DBA 任务，如数据库变更管理和数据库恢复。

数据库安全是另一个快速发展的领域，基于此需要对第 1 版进行重大调整。本书在安全功

---

⊖ 也称为“数据遮罩”。——编辑注



能和需求方面提供了全新的覆盖范围，包括基于标签的访问控制、加密和阻止 SQL 注入攻击。

本书增加了十年前尚未广泛采用的技术（如 XML），并且适当讨论了 DBA 应该关注的新生技术，包括 NoSQL 和云计算。同时，也覆盖了一些较新的 DBMS 功能，如时态数据库支持和 INSTEAD-OF 触发器。

最后，整本书经过了审查和修订以确保每个主题都添加了最新的技术和需求。本书特别关注那些 DBMS 功能用于凸显特定技术的例子是准确的并且是最新的。例如，DB2 HADR、SQL Server 2012 AlwaysOn 和 Oracle Transparent Data Encryption。

通过学习这本书，你可以得到实施和维护异构数据库管理方面最及时、最准确且最新的指导。你可以用它来了解成为一名成功的数据库管理员需要具备什么。还可以在将本书和特定 DBMS 产品供应商的手册或产品说明书结合使用。

## 如何使用这本书

这本书可以作为教程和参考。本书按照可能会遇到的所有 DBA 任务执行的时间顺序编排。因此，如果你从第 1 章到第 24 章依次阅读本书，你会对 DBA 的工作有一个全面、顺序的了解。你也可以单独阅读任何一章，因为每一章都涉及一个主题。如果书中其他材料对读者的理解有帮助，本书也明确指出了参考章节。

## 致 谢 *Acknowledgements*

写作是一项有益的工作，但是也需要大量的时间：调查、撰写、审查、编辑以及一遍又一遍地改写，直到满意为止。但没有人仅凭想象就可以写出一本技术类的书籍。在这本书的写作过程中，我遇到很多知识渊博且乐于助人的朋友，他们给予了太多的帮助。

首先，我想感谢许多业内专家审查了原书的提案。下面这些人对我的原作大纲提出了很多有用的建议和想法，使我的书变得更好，他们是：Michael Blaha、Keith W. Hare、Michael J. Hernandez、Robert S. Seiner 和 David L. Wells。此外，我要感谢大家在我开始写这本书之前肯花时间听我讲述构思。这份名单实在太长，我无法记住你们每一个人的名字，但你心里明白。

我想感谢那些审查全文并对其给予评论的人们。在本书的第 2 版中，Bill Arledge 和 Kevin Kline 在审查过程发挥他们的专长，提供了许多有益的建议并进行了大量修改，提高了本书的质量。我们不会忘记参与第 1 版审查的人员：Dan Hotka、Chris Foot、Chuck Kosin、David L. Wells 和 Anne Marie Smith，他们详细阅读了每章各种版本的稿件，这本书是对他们努力付出的回报。特别要感谢数据建模和管理大师 William J. Lewis 和 Robert S. Seiner，他们在百忙之中抽出时间对本书的第 3 章进行了审查，并提出了宝贵的建议。我还想感谢我的兄弟 Scott Mullins，他对第 5 章进行了审查，并对应用程序设计和开发提供了指导。

感谢 Mary Barnard，她出色地完成了第 1 版的编辑工作。还有 Greg Doench，同样出色地完成了第 2 版的编辑工作。她们两位使我的书更具可读性。

此外，感谢 Addison-Wesley 出版社那些与我一起工作的人，他们善解人意并且很有耐心，使得两个版本都能够顺利出版。他们是：Michelle Housley、Patrick Peterson、Stacie Parillo、Barbara Wood 和 Mary O'Brien，他们在协调书生产的整个过程中都给予了特别的帮助。

也要感谢我出色的妻子 Beth，有了她的理解和支持，我才可能写出这本书。事实上，要感谢我所有的家人和朋友一贯的支持和帮助。

最后，感谢与我有专业上合作的所有人，他们分别来自：SoftwareOnZ、NEON Enterprise Software、Embarcadero Technologies、BMC Software、Gartner Group、PLATINUM Technology, Inc.、Duquesne Light Company、Mellon Bank、USX Corporation 和 ASSET, Inc.。这本书之所以能变得更好，是因为我有幸与如此之多优秀的人员一起合作。

# 目 录 *Contents*

对本书的赞誉	
译者序	
前 言	
致 谢	
<b>第 1 章 什么是 DBA</b>	<b>1</b>
1.1 为什么学习数据库管理	2
1.2 独特的优势	3
1.2.1 DBA 的薪资	3
1.2.2 数据库技术	5
1.3 DBA 的管理准则	6
1.4 评估 DBA 的工作邀请	10
1.5 数据库管理、数据管理和系统管理	10
1.5.1 数据管理	10
1.5.2 数据库管理	13
1.5.3 系统管理	13
1.6 DBA 的任务	14
1.6.1 数据库设计	14
1.6.2 性能监控和调优	15
1.6.3 保证可用性	16
1.6.4 数据库安全和授权	16
1.6.5 治理与合规性	17
1.6.6 备份和恢复	18
1.6.7 确保数据完整性	18
1.7 DBMS 版本迁移	19
1.8 DBA 的类型	21
1.8.1 系统 DBA	21
1.8.2 数据库架构师	21
1.8.3 数据库分析师	22
1.8.4 数据建模师	22
1.8.5 应用程序 DBA	22
1.8.6 面向任务的 DBA	24
1.8.7 性能分析师	24
1.8.8 数据仓库管理员	24
1.9 人员配备的考虑	24
1.9.1 需要多少 DBA	25
1.9.2 DBA 报告结构	26
1.10 跨平台的 DBA 问题	28
1.11 生产环境与测试环境	29
1.12 新技术对 DBA 的影响	30
1.12.1 过程 DBA: 管理数据库逻辑	30
1.12.2 互联网: 从 DBA 到 eDBA	33

1.12.3 个人 DBA 与云计算 .....	35	2.3.6 DBA 人员技能组合 .....	60
1.12.4 NoSQL、大数据和 DBA .....	37	2.3.7 平台支持 .....	60
1.12.5 新技术对 DBA 的影响 .....	37	2.3.8 配套软件 .....	61
1.13 DBA 认证 .....	38	2.3.9 回滚计划 .....	61
1.14 本书的其余部分 .....	39	2.3.10 迁移验证 .....	62
1.15 回顾 .....	39	2.3.11 DBMS 升级策略 .....	62
<b>第 2 章 创建数据库环境</b> .....	<b>40</b>	2.4 数据库标准与过程 .....	62
2.1 定义企业的 DBMS 策略 .....	40	2.4.1 数据库命名约定 .....	62
2.1.1 DBMS 选型 .....	41	2.4.2 其他数据库标准和过程 .....	65
2.1.2 DBMS 架构 .....	45	2.5 DBMS 教育 .....	69
2.1.3 DBMS 集群 .....	47	2.6 小结 .....	70
2.1.4 DBMS 泛滥 .....	48	2.7 回顾 .....	70
2.1.5 硬件问题 .....	49	2.8 推荐书目 .....	71
2.1.6 云数据库系统 .....	49	<b>第 3 章 数据建模与规范化</b> .....	<b>72</b>
2.2 安装 DBMS .....	50	3.1 数据建模的概念 .....	73
2.2.1 DBMS 安装基础 .....	50	3.2 数据模型的组件 .....	77
2.2.2 硬件需求 .....	50	3.2.1 实体 .....	77
2.2.3 存储需求 .....	51	3.2.2 属性 .....	78
2.2.4 内存需求 .....	52	3.2.3 码 .....	82
2.2.5 配置 DBMS .....	53	3.2.4 关联 .....	82
2.2.6 连接 DBMS 与基础设施		3.3 发现实体、属性和关联 .....	84
配套软件 .....	54	3.4 概念、逻辑和物理数据模型 .....	85
2.2.7 安装验证 .....	54	3.5 什么是规范化 .....	87
2.2.8 DBMS 环境 .....	55	3.6 范式 .....	87
2.3 升级 DBMS 的主版本和次版本 .....	55	3.6.1 第一范式 .....	87
2.3.1 产品特性与复杂性 .....	58	3.6.2 第二范式 .....	89
2.3.2 DBMS 环境的复杂性 .....	59	3.6.3 第三范式 .....	90
2.3.3 DBMS 供应商的声誉 .....	59	3.6.4 一个规范化的数据模型 .....	91
2.3.4 DBMS 的支持条款 .....	59	3.6.5 进一步的范式 .....	91
2.3.5 企业风格 .....	60	3.7 规范化实践 .....	92

3.8 其他数据建模问题 .....	92	4.6 时态数据支持 .....	120
3.9 小结 .....	93	4.6.1 时态的例子 .....	120
3.10 回顾 .....	93	4.6.2 业务时间和系统时间 .....	120
3.11 推荐书目 .....	94	4.7 小结 .....	121
<b>第 4 章 数据库设计</b> .....	<b>96</b>	4.8 回顾 .....	122
4.1 从逻辑模型到物理数据库 .....	96	4.9 推荐书目 .....	123
4.1.1 实体转换成表 .....	97	<b>第 5 章 应用程序设计</b> .....	<b>124</b>
4.1.2 属性转换成列 .....	97	5.1 数据库应用程序开发和 SQL .....	124
4.1.3 为所有关联建立引用约束 .....	100	5.1.1 SQL .....	125
4.1.4 建立物理数据结构 .....	100	5.1.2 Set-at-a-Time 处理和 关系闭包 .....	127
4.2 数据库性能设计 .....	102	5.1.3 在程序中嵌入 SQL 语句 .....	128
4.2.1 索引设计 .....	102	5.1.4 SQL 中间件和 API .....	129
4.2.2 散列 .....	107	5.1.5 应用程序基础架构 .....	130
4.2.3 集群 .....	108	5.1.6 面向对象与 SQL .....	134
4.2.4 交叉数据 .....	109	5.1.7 SQL 类型 .....	135
4.3 非规范化 .....	109	5.1.8 SQL 性能编码 .....	136
4.3.1 何时进行非规范化 .....	109	5.1.9 查询 XML 数据 .....	137
4.3.2 预连接表 .....	111	5.2 定义事务 .....	138
4.3.3 报告表 .....	111	5.2.1 事务指南 .....	139
4.3.4 镜像表 .....	112	5.2.2 工作单元 .....	140
4.3.5 拆分表 .....	112	5.2.3 事务处理系统 .....	140
4.3.6 组合表 .....	114	5.2.4 应用程序服务器 .....	141
4.3.7 冗余数据 .....	114	5.3 锁 .....	142
4.3.8 重复组 .....	114	5.3.1 锁的类型 .....	143
4.3.9 可导性数据 .....	115	5.3.2 锁超时 .....	144
4.3.10 层次结构 .....	116	5.3.3 死锁 .....	144
4.3.11 特殊物理实施需求 .....	117	5.3.4 锁定持续时间 .....	145
4.3.12 非规范化小结 .....	117	5.3.5 锁升级 .....	148
4.4 视图 .....	118	5.3.6 最小化锁定问题的编程技术 .....	148
4.5 数据定义语言 .....	119		

5.3.7 锁定小结 .....	149	7.2.1 DBMS 软件 .....	167
5.4 批处理 .....	149	7.2.2 硬件配置 .....	167
5.5 小结 .....	150	7.2.3 逻辑和物理设计 .....	167
5.6 回顾 .....	150	7.2.4 应用程序 .....	168
5.7 推荐书目 .....	151	7.2.5 物理数据库结构 .....	168
<b>第 6 章 设计审查</b> .....	<b>153</b>	7.3 数据库结构变更的影响 .....	169
6.1 什么是设计审查 .....	153	7.3.1 ALTER 的限制 .....	170
6.1.1 交战规则 .....	154	7.3.2 数据库变更业务场景 .....	171
6.1.2 设计审查的参与者 .....	154	7.3.3 比较数据库结构 .....	173
6.1.3 所需的知识与技能 .....	156	7.3.4 请求数据库变更 .....	174
6.2 设计审查的类型 .....	157	7.3.5 标准化的变更请求 .....	174
6.2.1 概念设计审查 .....	158	7.3.6 沟通 .....	175
6.2.2 逻辑设计审查 .....	158	7.3.7 协调数据库和应用程序变更 .....	175
6.2.3 物理设计审查 .....	159	7.3.8 合规性 .....	176
6.2.4 组织设计审查 .....	159	7.3.9 DBA 脚本和变更管理 .....	176
6.2.5 SQL 与应用程序代码设计 审查 .....	160	7.4 小结 .....	177
6.2.6 部署前的设计审查 .....	160	7.5 回顾 .....	177
6.2.7 部署后的设计审查 .....	161	7.6 推荐书目 .....	177
6.3 设计审查输出 .....	161	<b>第 8 章 数据可用性</b> .....	<b>178</b>
6.4 其他注意事项 .....	161	8.1 定义可用性 .....	179
6.4.1 应对远程工作人员 .....	162	8.2 停机成本 .....	182
6.4.2 制度和知识迁移 .....	162	8.3 可用性问题 .....	184
6.5 小结 .....	162	8.3.1 数据中心的丢失 .....	184
6.6 回顾 .....	162	8.3.2 网络问题 .....	185
6.7 推荐书目 .....	163	8.3.3 服务器硬件的丢失 .....	185
<b>第 7 章 数据库变更管理</b> .....	<b>164</b>	8.3.4 磁盘相关的中断 .....	187
7.1 变更管理要求 .....	165	8.3.5 操作系统故障 .....	187
7.2 变更的类型 .....	167	8.3.6 DBMS 软件故障 .....	188
		8.3.7 应用程序问题 .....	188
		8.3.8 安全和授权问题 .....	188

8.3.9 数据损坏 .....	189	9.6 DBMS 的性能基础 .....	212
8.3.10 数据库对象丢失 .....	189	9.7 小结 .....	213
8.3.11 数据丢失 .....	190	9.8 回顾 .....	213
8.3.12 数据复制和传播故障 .....	190	9.9 推荐书目 .....	214
8.3.13 严重的性能问题 .....	191	<b>第 10 章 系统性能</b> .....	215
8.3.14 恢复问题 .....	191	10.1 大环境 .....	215
8.3.15 DBA 错误 .....	191	10.1.1 与操作系统交互 .....	216
8.3.16 计划内和计划外中断 .....	193	10.1.2 联合代理 .....	216
<b>8.4 确保可用性</b> .....	193	10.1.3 硬件配置 .....	217
8.4.1 系统正常运行时执行日常 维护 .....	194	10.1.4 DBMS 组件 .....	218
8.4.2 自动化 DBA 任务 .....	195	<b>10.2 DBMS 安装和配置问题</b> .....	220
8.4.3 利用高可用性功能 .....	196	10.2.1 配置类型 .....	221
8.4.4 利用集群技术 .....	196	10.2.2 内存使用 .....	221
8.4.5 数据库架构和 NoSQL .....	199	10.2.3 数据高速缓存 .....	224
<b>8.5 小结</b> .....	200	10.2.4 “开放的”数据对象 .....	227
<b>8.6 回顾</b> .....	200	10.2.5 数据库日志 .....	227
<b>8.7 推荐书目</b> .....	201	10.2.6 锁定与争用 .....	231
<b>第 9 章 性能管理</b> .....	202	10.2.7 系统目录 .....	231
9.1 性能定义 .....	202	10.2.8 其他配置选项 .....	232
9.2 监测与管理 .....	205	10.2.9 一般性建议 .....	233
9.2.1 被动与主动 .....	206	<b>10.3 系统监测</b> .....	233
9.2.2 生产前的性能评估 .....	207	<b>10.4 小结</b> .....	234
9.2.3 历史趋势 .....	208	<b>10.5 回顾</b> .....	234
<b>9.3 服务级别管理</b> .....	208	<b>10.6 推荐书目</b> .....	235
<b>9.4 性能调优的类型</b> .....	210	<b>第 11 章 数据库性能</b> .....	236
9.4.1 系统调优 .....	210	11.1 优化数据库的技术 .....	236
9.4.2 数据库调优 .....	210	11.1.1 分区 .....	237
9.4.3 应用程序调优 .....	211	11.1.2 原始分区与文件系统 .....	237
<b>9.5 性能调优工具</b> .....	211	11.1.3 索引 .....	238



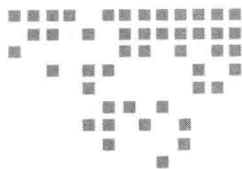
11.1.4 非规范化 .....	240	12.6 小结 .....	278
11.1.5 集群 .....	241	12.7 回顾 .....	278
11.1.6 交叉数据 .....	243	12.8 推荐书目 .....	278
11.1.7 可用空间 .....	243		
11.1.8 压缩 .....	244	<b>第 13 章 数据完整性</b> .....	280
11.1.9 文件放置与分配 .....	245	13.1 完整性的类型 .....	280
11.1.10 页面大小 (块大小) .....	246	13.2 数据库结构完整性 .....	281
11.2 数据库重组 .....	247	13.2.1 结构性问题的类型 .....	281
11.2.1 决定何时进行重组 .....	250	13.2.2 管理结构性问题 .....	281
11.2.2 自动化 .....	251	13.3 语义数据的完整性 .....	284
11.3 小结 .....	251	13.3.1 实体完整性 .....	285
11.4 回顾 .....	251	13.3.2 唯一约束 .....	286
11.5 推荐书目 .....	252	13.3.3 数据类型 .....	286
		13.3.4 默认值 .....	287
<b>第 12 章 应用程序性能</b> .....	253	13.3.5 检查约束 .....	287
12.1 设计关系访问的应用程序 .....	253	13.3.6 触发器 .....	292
12.2 关系优化 .....	254	13.3.7 参照完整性 .....	296
12.2.1 CPU 和 I/O 成本 .....	255	13.4 时态数据库系统 .....	304
12.2.2 数据库统计信息 .....	255	13.5 小结 .....	305
12.2.3 查询分析 .....	257	13.6 回顾 .....	306
12.2.4 连接 .....	257	13.7 推荐书目 .....	306
12.2.5 访问路径选择 .....	259		
12.3 其他优化考虑 .....	266	<b>第 14 章 数据库安全</b> .....	308
12.3.1 视图访问 .....	266	14.1 数据泄露 .....	308
12.3.2 查询重写 .....	267	14.2 数据库安全基础 .....	309
12.3.3 基于规则的优化 .....	267	14.3 授予和撤销授权 .....	313
12.4 访问路径的审查 .....	268	14.3.1 权限类型 .....	313
12.5 SQL 的编码和调优效率 .....	272	14.3.2 授予 PUBLIC .....	315
12.5.1 SQL 经验法则—箩筐 .....	272	14.3.3 撤销权限 .....	316
12.5.2 其他 SQL 调优技巧 .....	277	14.3.4 基于标签的访问控制 .....	317
12.5.3 识别性能低下的 SQL .....	277	14.3.5 安全报告 .....	319

14.4 授权角色和组 .....	319	15.5 用于长期数据保留的数据库 归档 .....	342
14.4.1 角色 .....	319	15.5.1 数据的生命周期 .....	342
14.4.2 组 .....	320	15.5.2 数据库归档 .....	343
14.5 其他数据库安全机制 .....	321	15.5.3 数据库归档解决方案的 构成 .....	346
14.5.1 安全使用视图 .....	321	15.5.4 e-Discovery 对 DBA 的影响 .....	347
14.5.2 安全使用存储过程 .....	322	15.6 密切跟踪传统的 DBA 任务 .....	348
14.6 加密 .....	322	15.6.1 数据库变更管理 .....	348
14.6.1 静态数据加密 .....	323	15.6.2 数据库备份和恢复 .....	349
14.6.2 传输过程的数据加密 .....	324	15.7 小结 .....	350
14.6.3 加密技术 .....	324	15.8 回顾 .....	351
14.7 SQL 注入攻击 .....	325	15.9 推荐书目 .....	351
14.8 审计 .....	327		
14.9 外部安全 .....	328	<b>第 16 章 数据库备份和恢复 .....</b>	<b>353</b>
14.9.1 作业调度和安全 .....	329	16.1 备份和恢复的重要性 .....	353
14.9.2 非 DBMS 的 DBA 安全 .....	329	16.2 问题的准备工作 .....	354
14.10 DBMS 补丁包和维护 .....	330	16.3 备份 .....	354
14.11 小结 .....	330	16.3.1 完整备份与增量备份 .....	357
14.12 回顾 .....	330	16.3.2 数据库对象和备份 .....	358
14.13 推荐书目 .....	331	16.3.3 DBMS 控制 .....	359
<b>第 15 章 合规性和数据库管理 .....</b>	<b>332</b>	16.3.4 并发访问问题 .....	360
15.1 实现合规性的协作方法 .....	334	16.3.5 备份的一致性 .....	361
15.2 元数据管理、数据质量和 数据治理 .....	335	16.3.6 日志归档和备份 .....	362
15.2.1 元数据 .....	335	16.3.7 确定备份计划 .....	363
15.2.2 数据质量 .....	336	16.3.8 DBMS 实例备份 .....	365
15.2.3 数据治理 .....	336	16.3.9 设计恢复的 DBMS 环境 .....	365
15.3 数据库审计和数据访问跟踪 .....	336	16.3.10 数据库备份的替代途径 .....	365
15.3.1 数据库审计技术 .....	338	16.3.11 记录备份策略 .....	367
15.3.2 特权用户审计 .....	340	16.3.12 数据库对象定义备份 .....	367
15.4 数据屏蔽和模糊 .....	341	16.4 恢复 .....	368
		16.4.1 确定恢复选项 .....	368

16.4.2 数据库对象恢复的一般步骤 .....	370	<b>第 18 章 数据和存储管理</b> .....	396
16.4.3 恢复的类型 .....	371	18.1 存储管理基础 .....	396
16.4.4 索引恢复 .....	377	18.2 文件和数据集 .....	399
16.4.5 测试恢复计划 .....	377	18.2.1 磁盘上的文件放置 .....	399
16.4.6 恢复删除的数据库对象 .....	378	18.2.2 原始分区和文件系统 .....	401
16.4.7 恢复损坏的块和页面 .....	379	18.2.3 临时数据库文件 .....	401
16.4.8 填充测试数据库 .....	379	18.3 空间管理 .....	401
16.5 替代性备份和恢复 .....	379	18.3.1 数据页面布局 .....	402
16.5.1 备用数据库 .....	380	18.3.2 索引页面布局 .....	404
16.5.2 复制 .....	380	18.3.3 事务日志 .....	406
16.5.3 磁盘镜像 .....	381	18.4 碎片和存储 .....	407
16.6 小结 .....	381	18.5 存储选项 .....	407
16.7 回顾 .....	381	18.5.1 RAID .....	408
16.8 推荐书目 .....	382	18.5.2 JBOD .....	414
<b>第 17 章 防灾规划</b> .....	383	18.5.3 存储区域网络 .....	414
17.1 计划的必要性 .....	383	18.5.4 网络附加存储 .....	414
17.2 一般灾难恢复指南 .....	386	18.5.5 分层存储 .....	415
17.2.1 远程站点 .....	386	18.6 规划未来 .....	416
17.2.2 书面计划 .....	386	18.7 小结 .....	417
17.2.3 人员 .....	389	18.8 回顾 .....	417
17.3 为灾难恢复备份数据库 .....	389	18.9 推荐书目 .....	418
17.3.1 磁带备份 .....	390	<b>第 19 章 数据移动与分布</b> .....	419
17.3.2 存储管理备份 .....	391	19.1 装卸数据 .....	419
17.3.3 其他方法 .....	392	19.1.1 LOAD 实用程序 .....	420
17.3.4 一些指导 .....	392	19.1.2 UNLOAD 实用程序 .....	422
17.4 灾难防治 .....	393	19.1.3 维护应用程序测试台 .....	425
17.5 小结 .....	394	19.2 导出与导入 .....	426
17.6 回顾 .....	394	19.3 批量数据移动 .....	426
17.7 推荐书目 .....	395	19.3.1 ETL 软件 .....	426
		19.3.2 复制和传播 .....	427

19.3.3 消息软件 .....	427	<b>第 21 章 数据库连接 .....</b>	<b>452</b>
19.3.4 其他方法 .....	428	21.1 多层、分布式计算 .....	452
<b>19.4 分布式数据库 .....</b>	<b>428</b>	21.1.1 回顾历史 .....	452
19.4.1 建立分布式环境 .....	430	21.1.2 业务问题 .....	453
19.4.2 数据分布的标准 .....	431	21.1.3 C/S 计算简介 .....	453
19.4.3 访问分布式数据 .....	432	21.1.4 C/S 应用程序的类型 .....	456
19.4.4 两阶段 COMMIT .....	433	<b>21.2 网络流量 .....</b>	<b>459</b>
19.4.5 分布式性能问题 .....	433	21.2.1 数据库网关 .....	459
<b>19.5 小结 .....</b>	<b>434</b>	21.2.2 数据库驱动程序 .....	460
<b>19.6 回顾 .....</b>	<b>434</b>	21.2.3 连接池 .....	461
<b>19.7 推荐书目 .....</b>	<b>435</b>	<b>21.3 数据库、互联网和 Web .....</b>	<b>461</b>
<b>第 20 章 数据仓库管理 .....</b>	<b>436</b>	21.3.1 接入互联网的数据库 .....	462
20.1 什么是数据仓库 .....	436	21.3.2 Web 开发和 Web 服务 .....	465
20.2 管理数据仓库 .....	438	<b>21.4 小结 .....</b>	<b>466</b>
20.2.1 太专注于技术 .....	439	<b>21.5 回顾 .....</b>	<b>466</b>
20.2.2 数据仓库设计 .....	439	<b>21.6 推荐书目 .....</b>	<b>467</b>
20.2.3 数据移动 .....	441	<b>第 22 章 元数据管理 .....</b>	<b>469</b>
20.2.4 数据清理 .....	441	22.1 元数据简介 .....	469
20.2.5 数据仓库的可扩展性 .....	444	22.1.1 从数据到知识, 再到更多 .....	470
20.2.6 数据仓库的性能 .....	444	22.1.2 元数据策略 .....	471
20.2.7 数据新鲜度 .....	447	22.1.3 数据仓库和元数据 .....	471
20.2.8 数据内容 .....	447	<b>22.2 元数据的类型 .....</b>	<b>471</b>
20.2.9 数据使用 .....	448	<b>22.3 资源库和数据字典 .....</b>	<b>473</b>
20.2.10 财务退款 .....	448	22.3.1 资源库的好处 .....	474
20.2.11 备份与恢复 .....	449	22.3.2 资源库的挑战 .....	475
20.2.12 不要凭空操作 .....	449	22.3.3 数据字典 .....	476
<b>20.3 小结 .....</b>	<b>450</b>	<b>22.4 小结 .....</b>	<b>476</b>
<b>20.4 回顾 .....</b>	<b>450</b>	<b>22.5 回顾 .....</b>	<b>477</b>
<b>20.5 推荐书目 .....</b>	<b>450</b>	<b>22.6 推荐书目 .....</b>	<b>477</b>

<b>第 23 章 DBA 工具</b> .....	479	24.3 自动化	505
23.1 DBA 工具的类型和好处 .....	479	24.4 知识共享	507
23.1.1 数据建模和设计 .....	479	24.5 分析、简化和专注	508
23.1.2 数据库变更管理 .....	480	24.6 不要惊慌	509
23.1.3 表编辑器 .....	484	24.7 测量两次，切割一次	509
23.1.4 性能管理 .....	485	24.8 了解业务，不只是技术	510
23.1.5 备份与恢复 .....	489	24.9 不要成为隐士	511
23.1.6 数据库实用程序 .....	490	24.10 任意使用所有的资源	511
23.1.7 数据保护、治理、风险和 合规性工具 .....	491	24.11 不断更新	512
23.1.8 数据仓库、分析和商业 智能 .....	494	24.12 投资自己	512
23.1.9 编程和开发工具 .....	496	24.13 小结	513
23.1.10 杂项工具 .....	497	24.14 最后的检验	513
23.2 研究原生 DBA 工具 .....	499	<b>附录 A 数据库基本原理</b> .....	516
23.3 评估 DBA 工具供应商 .....	500	<b>附录 B DBMS 供应商</b> .....	521
23.4 小结 .....	502	<b>附录 C DBA 工具供应商</b> .....	526
23.5 回顾 .....	503	<b>附录 D DBA 的网络资源</b> .....	531
<b>第 24 章 DBA 的经验法则</b> .....	504	<b>附录 E DBA 公开招聘样本</b> .....	538
24.1 记录每一件事 .....	504	<b>参考书目</b> .....	542
24.2 一切都井然有序 .....	504	<b>术语表</b> .....	563



# 什么是 DBA

每一家使用数据库管理系统 (DBMS) 管理数据的公司都需要数据库管理 (DBA) 组来确保能够有效地使用和部署公司的数据库。如今各种规模的企业都会至少使用一种 DBMS，这使得对数据库管理员 (DBA) 的需求比以往任何时候都要多。然而，DBA 的准则要么不容易理解，要么在推广时不能使用。

对数据库管理员 (DBA) 的需求比以往任何时候都要多。

关于数据库管理，有个经常说起的笑话，它可以帮助我们认识 DBA 的必要性和我们对 DBA 工作认知的不足。笑话大概是这样的：

---

Acme 公司的 CIO 聘请一家管理顾问公司来帮助他们简化工作流程。为了了解 Acme 的工作方式，顾问与 CIO 进行了面谈。其中一个问题是：“嗯，我看到你们员工中有 DBA，他是干什么的？”

CIO 回答道：“哦，我们使用 Oracle，有人跟我说，我们必须要用 DBA 来保证 Oracle 数据库长期在线。我知道，我们的一些关键业务流程如订单录入和库存会使用 Oracle，我还真不知道 DBA 是干什么的。” CIO 继续说，“请不要告诉我还要一个吧，就这个的工资我们都付不起了！”

---

悲哀吧，但这往往是事实，许多公司的数据库管理就是这样的状况。DBA 经常会被视作一个牧师或者魔术师，他们用魔术使得数据库系统能够有效运作。如今的 DBMS 软件非常复杂，导致只有极少数人能够像了解基础知识（如 SQL）那样了解它们。但是 DBA 只有了解了 DBMS 的复杂性，才能变得更有价值。事实上，很多时候公司内既懂数据库管理又懂开发

知识的只有 DBA。

DBA 的角色就是公司数据资产的监护人。所以，DBA 会小心翼翼地保护着他们的数据。而开发人员负责开发新的应用程序，他们会经常遇到挑战，要求快速开发，一个项目接着一个项目。很明显，两种角色和他们对工作的期望的不同，会在他们之间制造出摩擦——DBA 说我要“变更控制，变更管理”，而开发人员则说我要“现在就部署”。

对 DBA 常有的另一个评价是他们很难相处。他们时常被视作自命不凡，拥有丰富的技术知识但人际交往能力有限。大约每一个数据库程序员都有个关于 DBA 的故事。这些趣闻轶事经常会以“我有个问题请教……”开头，而以“……最后他（DBA）告诉我不要打扰他了，自己去看说明书”结束。DBA 在人们心中的形象不是“暖融融”的，大部分是他们的工作特点和范围所致。DBMS 贯穿了整个企业，企业需要制订有效使用 DBA 的流程。

事实上，DBA 经常不得不坐下来独自完成一些事情，这或许能够缓解他们的负面形象。许多数据库问题需要一段时间来安静地思考和分析才能够解决。所以，DBA 一般不喜欢被打扰。但是，即便有一些问题需要他们独自解决，还是会有许多的问题需要团队合作。而且由于大部分的 DBA 拥有丰富的知识，实际上，他们的工作经常会被回答和解决问题中断，因此他们安静思考的时间会很少。

DBA 的孤立行为不应该得到肯定，而应该通过培训使他们获得良好的沟通技巧。数据是应用程序的命脉。开发应用程序就是对数据的读写、分析、移动、计算和修改等。没有数据，程序就一无是处。DBA 处在整个开发生命周期的核心位置——他们确保应用程序能够有效、准确地访问企业数据。通常，DBA 经常和许多不同种类的人打交道，包括技术人员、程序员、终端用户、客户和领导。然而，许多 DBA 过于关注 DBMS 的内部技术细节，而忽略了提高与他们同事或客户的沟通技巧。

DBA 需要学习良好的沟通技巧。

到现在，我们还没有回答本章标题的问题：什么是 DBA？简短地回答：DBA 是信息方面的专家，他们确保企业的日常工作正常运转，确保企业的数据库以及访问这些数据库的应用程序的有效性。

DBA 是信息方面的专家，他们确保企业的日常工作正常运转，确保企业的数据库以及访问这些数据库的应用程序的有效性。

而本书，就是回答这个问题的具体答案。本书将会定义数据库管理的管理准则，并且为 DBA 的工作顺利进行提供练习指导。为了开始回答这个问题，本章将会提供数据库管理的概括性介绍，包括为什么把 DBA 作为一种职业？DBA 深层次工作介绍，还有企业 DBA 团队的组织架构。

## 1.1 为什么学习数据库管理

我们已经提到过，数据是应用程序的核心，如今现代化公司离开了数据就不能运行。从

某种意义上来说,现如今业务就是数据。<sup>①</sup>如果没有数据,企业将没有能力来管理财政、进行交易,或联系客户。创建数据库就是为了存储并组织这些数据。数据库的设计越好,对它们的利用越好,公司的商业竞争力会越强。

诸多IT公司面临的最大问题之一就是怎样确保数据库管理的质量。根据最近的一项调查显示:<sup>②</sup>

- 好的DBA很难找,而且要价很高——76%的受调查对象表示最少要用3个月才能招到一名DBA,就算培训一名新的DBA也要3个月以上。
- 支撑应用程序的数据库基础设施复杂,长期分散,且不易于管理。

所有这些调查结果都清晰地表明数据库管理是一项有难度的、要求非常高的工作。还有,根据Dice<sup>③</sup> 2010-2011的技术薪资调查表明,每天发布的工作需求中有超过15 000份工作要求有Oracle经验。对Oracle技术的需求每年都高达57%,而有Oracle工作经验的技术专家的平均薪资是90 914美元。

## 1.2 独特的优势

一名优秀的DBA要享受挑战并且还得是出色的问题终结者。

DBA负责设计和维护企业的数据库,他处在企业的核心位置。这样,DBA就有机会去学习各个方面的业务,以及知晓知识间的关联。他们还能研究公司的前沿技术,这使得他们的工作充满了新鲜感——但是第一次尝试找出一项新技术是怎样工作的过程中可能会有挫折感。DBA常常是独自努力研究,在遇到问题的时候不会有其他专家帮忙。因此,一名优秀的DBA要享受挑战并且还得是出色的问题终结者。

### 1.2.1 DBA的薪资

作为技术人员你不会找到比DBA更有挑战的IT相关工作了。幸运的是,DBA的工作报酬也是相当丰厚的。根据一项由全球知识(Global Knowledge)<sup>④</sup>与技术团体组织的薪资调查,<sup>⑤</sup>DBA的薪资是78 468美元,而DBA管理者的平均薪资是87 261美元。这些平均薪资的数字比另外一项Janco<sup>⑥</sup>薪资调查(前些年进行的)略低。

对于全职的DBA来说,既定薪资最高可达80 000美元。表1-1详细给出了在2006~2011

① 社交媒体使得这种趋势更加明显。人们将自己正在经历的生活以数据驱动的形式记录下来,并向世界大声宣扬着。

② 2006年题为“数据库管理的趋势:自动化机遇”的GridApp系统研究。

③ Dice.com是一个求职网站,是技术和工程专业人员,以及需要这些人员的公司的一个平台。

④ 全球知识是一家技术培训机构。

⑤ Tech Republic是ZDNet公布的门户网站。

⑥ Janco Associates, Inc.是一家致力于管理信息系统的管理咨询公司。



年间 DBA 的既定薪资（数据来自 Dice 2010-2011 的技术薪资调查）。

表 1-1 2006 到 2011 年 DBA 的既定薪资

年 份	DBA 的既定薪资
2006 ~ 2007	85 441 美元
2007 ~ 2008	85 092 美元
2008 ~ 2009	89 742 美元
2009 ~ 2010	91 283 美元
2010 ~ 2011	88 443 美元

请注意这里引用的薪资数字仅作说明之用，它们会随多种因素而变化。随着工作年限和从业人数的增加，薪资也必然随之增加。当然了，DBA 的薪资和其他人的薪资一样也会因地域的不同而不同。在美国，东北部地区和西海岸的 DBA 薪资有可能是比其他区域高。行业类型也是因素之一，比如，制药厂比政府部门的薪资要高。当这些所有的因素都考虑到了，DBA 的薪资数字就会稳居 IT 从业人员之首。

根据 2012 年计算机世界杂志的薪资调查，全国 DBA 的平均薪资是 95 187 美元，而一个有着 15 ~ 20 年工作经验且在太平洋海岸工作的 DBA 的平均薪资却是 103 597 美元。<sup>①</sup>显然，工作经验和工作地点造成了显著的差异。

美国劳工统计局（BLS）<sup>②</sup>提供了有关 DBA 雇佣和报酬的附加信息。自 2012 年 5 月起，BLS 的报告称数据库管理员的年平均工资的中线是 73 490 美元，而既定年平均工资是 75 730 美元。10% 的低工资的人挣得少于 41 570 美元，而 10% 的高工资的人挣得高于 115 660 美元。BLS 还对收入统计按照地理区域进行了划分，你还可以根据他们的数据来确定你所在区域的期望薪资范围。

或许你的再就业能力比报酬更重要，受雇作为一名 DBA 的前景是相当好的。根据美国劳工统计局职业展望手册 2010—2011 版，“预计就业增长速度超过了平均值，工作的前景应当是非常好的”。事实上，在 2008 ~ 2018 年间，BLS 预估 DBA 工作的数量会增加 20%。

DBA 的工作市场增长速度超过了平均值。

所以，DBA 是高薪资、高就业率、极具挑战性的工作，并且有可能从事最紧急、最重要的项目。有什么理由不喜欢吗？是的，人们期望 DBA 知道一切，不只是数据库技术，还包括与之相关的一切。DBA 几乎从来没有每天只工作 8 小时，经常是连续好几天加班，特别是当数据库的性能出现问题或者开发中的项目落后于预定计划时。据行业分析师分析，DBA 的每周平均工作时间是 50+ 小时，其中包含周末 6 小时。DBA 经常不得不在周末和节假日这些数据库使用的低谷时间去维护数据库。

① [www.computerworld.com/s/article/9224243/IT\\_Salary\\_Survey\\_2012](http://www.computerworld.com/s/article/9224243/IT_Salary_Survey_2012)。

② [www.bls.gov/](http://www.bls.gov/)。

数据库管理是一项没有休息的工作。

DBA 不但是一项技术上有挑战且有回报的工作，还是一项有可能使你筋疲力尽、饱受挫折的工作。但是千万不要让这些把你吓住了，这项工作的好的方面远比不好的方面多。

### 1.2.2 数据库技术

DBA 是精通数据库技术具体细节的 IT 专业人士。因此，他们对数据库管理基本原理有扎实的理解非常重要。这就要求不仅要懂得行、列和表格或 SQL 这些简单的知识了。

本书假设读者对关系数据库技术和 DBMS 基本原理已有基本的了解。对于需要回顾这些概念的读者，请参阅附录 A。“数据库基础”可不是一件小事。问题是，有时人们认为他们知道的很多。例如，这个问题：什么是数据库？我敢说大多数的人相信他们知道这个问题的答案。但其中一部分人（或者很多）都会答错。SQL Server 不是数据库，它只是一种 DBMS（数据库管理系统）。你可以使用 SQL Server 来创建数据库，但 SQL Server 本身不是一种数据库。

那么，什么是数据库？数据库是一种有组织的数据存储，其中的数据可以通过指定的数据元素（比如，字段、记录和文件）来访问（详见图 1-1）。

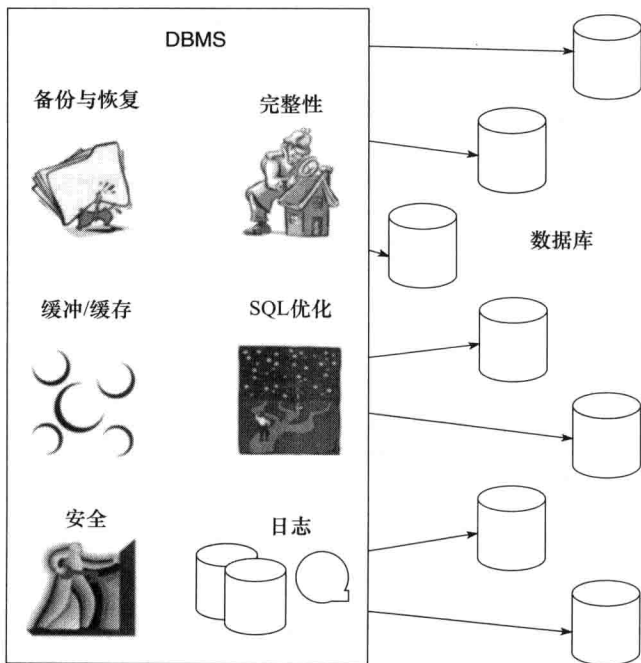


图 1-1 DBMS 对数据库

数据库是一种有组织的数据存储，其中的数据可以通过指定的数据元素来访问。

DBMS 是一种使终端用户或程序员能够共享数据的软件。它提供了一套操作数据库的系

统性方法：创建、更新、检索和存储信息。DBMS 通常还负责数据完整性、数据安全性、数据访问控制和优化、自动回退、重起和恢复。

对外行人来说，你可以把数据库想象成一个文件夹，把 DBMS 想象成文件柜，且所容纳的所有文件都贴有标签。DBMS 管理着数据库，通过 DBMS 来实现和访问数据库实例。所以说，DB2、Oracle 和 SQL Server 都是数据库管理系统。工资单应用使用工资单的数据库，该数据库可以使用 DB2、Oracle 或者 SQL Server 来实现。

为什么 DBMS 如此重要呢？因为如果在工作区使用的术语不精确，就会引起混淆，进而导致项目超出预算，开发的系统不正常，失去生产力。

除了数据库管理的基本原理，DBA 还要对使用的所有 DBMS 产品（一家企业可能使用多种）非常精通。例如，一家大型的企业可能在中央主机使用 DB2，而在几个不同的 UNIX 平台上使用 Oracle 和 Informix，在 Linux 上使用 MySQL，在 Windows 上使用 SQL Server。旧的遗留系统上可能使用的是 IMS 数据库，然而的确有个别“疯狂的程序”使用了边缘的 DBMS，如 Adabas 或 Ingres。<sup>①</sup>还有一些新的数据库技术，如 NoSQL，以及列存储的 DBMS 产品（如 Hadoop）。还有云数据库系统，如 Microsoft SQL Azure 和 Google BigTable。

因此，DBA 组要有应对这些不同的数据库管理系统和平台的专业知识。与此同时，DBA 还要能够确定哪些数据库管理系统和平台最能满足某个应用程序的需要。这是一项艰难的工作，充满了博弈和冲突意见。这时候 DBA 组要通过对比每种数据库管理系统和平台，然后根据应用程序的需要来做出正确的选择。

DBA 要决定最匹配的应用程序、DBMS 和平台。

有关 DBMS 概念的简短介绍，请参阅附录 A。

## 1.3 DBA 的管理准则

数据库管理很少被视作一种管理准则。“准则”一词意味着规划并按照该规划实施。当数据库管理被视作一项管理准则时，公司内部的数据处理就会有所改善了。这就是消极被动和积极主动的区别。

DBA 组让需求和问题淹没是家常便饭。这有多种原因，包括人员缺乏、过度承诺支持新的（甚至现有的）应用程序开发项目、缺少可遵循的流程和缺少预算等。这种情况下，数据库管理员就会变得消极被动。消极被动的 DBA 更像是名消防员，他集中注意力去解决所关注的最大问题。换句话说，消极被动的 DBA 只有当问题发生后才尝试去解决它们。

积极主动的 DBA 在企业内部部署数据库时，通过制订并执行战略蓝图，以避免诸多问题。该蓝图应当针对应用程序开发周期（ADLC，详见图 1-2）的所有阶段。积极主动的

---

① 之所以将 Adabas 和 Ingres 视为“边缘”，并不是因为任何功能和技术缺陷，而是它们所占的市场份额较小。

DBA 通过执行规范和操作流程，避免问题发生。

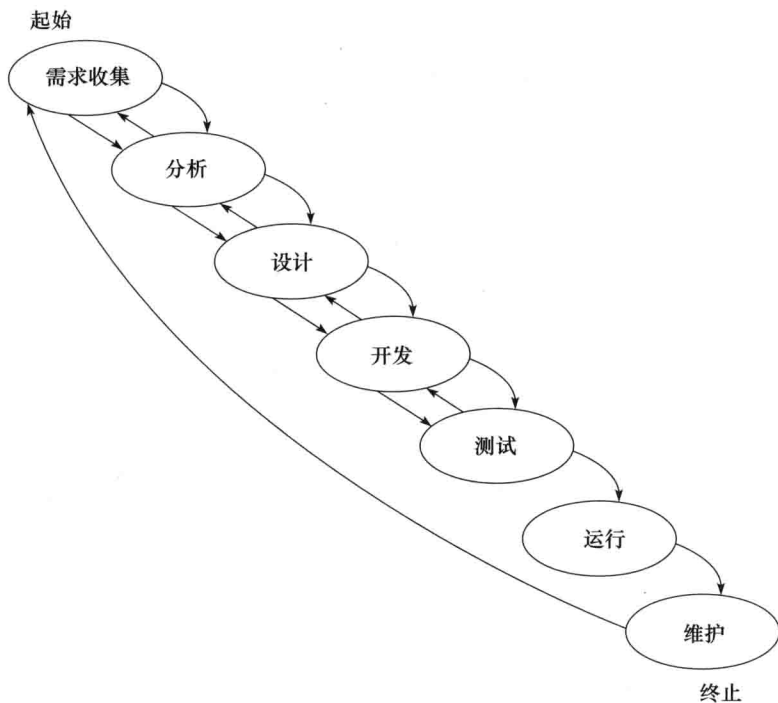


图 1-2 应用程序开发生命周期 (ADLC)

积极主动的 DBA 能避免许多问题。

数据专家，通常指 DBA，应当参与 ADLC 的每一个阶段。在项目启动和需求收集阶段，DBA 就要着手识别出项目的数据成分。他能够帮助确认所需的数据已在别处存在或者是崭新的。在分析和设计阶段，他要把基本数据需求转换成一种概念性的、逻辑上的数据模块。

开始开发前，逻辑上的数据模块要转换为一种物理的数据库设计，它可以用 DBMS 实施，如 Oracle 或 DB2。同时，样本数据必须要植入物理数据库中，用于测试应用程序。除此之外，DBA 还要通过开发并实施一种流程去刷新测试数据，进而使得重复的测试能够进行。

当应用程序开发完毕并开始运行时，DBA 要确保 DBMS 为新负载做好了准备。包括：执行适当的安全措施，衡量并修改新应用程序对存储盘和内存的需求，预估新负载对现有的数据库和应用程序会带来什么影响。DBA 也负责新数据库从测试环境到生产环境的迁移。

应用程序运行时，DBA 要履行“主机”的职责，包括性能和有效性监控、调试、备份和恢复、权限管理。但没有程序或数据库是长期保持不变的，因为业务的需求会改变，支撑业务的 IT 系统也会改变。当需要维护时，DBA 又要再次忙于从需求收集到应用程序运行的整个过程了。当监控发现了一处性能缺陷时，DBA 要向开发团队（如果问题出在程序代码上），

或者向负责存储的团队（如果 I/O 是瓶颈）建议更好的性能方案，或者推荐其他的解决方案，如新的或提升了的标准、替代性的 SQL 代码，等等。

## 隐私条款与隐私数据

数据隐私是一个新兴的问题，正在成为越来越多的企业的负担。不仅要管理的数据量增加了，新数据出现的速度也增加了。

在这个网络化的世界中，我们彻底数字化了。我们有 ID、位置、操作、采购、关联、动作，还有存储的历史数据。我们不禁要问了：是谁在收集这些数据？他们要用这些数据干什么？他们和谁共享这些数据？最重要的是，每个人都会问的是“我怎样保护我的信息不被滥用”。这些都是合情合理的问题，我们都想知道答案。

在美国，有超过 30 种联邦法规和超过 100 种国家法规来管制信息隐私，而对个人隐私却一直是采用头痛医头、脚痛医脚的方式。另一方面，欧洲联盟已经下达了数据保护指令，要求其成员国通过法律实现其职权。该指令为信息收集的受众赋予了权利，称受众为“数据当事人”。收集信息的工作人员必须要向数据当事人说明谁在收集数据？最终谁能够访问这些数据？为什么要收集数据？数据当事人也有权限访问和修改他们的数据。

甚至还有数据隐私日，<sup>①</sup>其目的是为了表达对那些通过个人信息表达自己的人的尊重。

最近破产的 Borders 书店提供了一个隐私政策影响企业数据的例子。Borders 停止经营的同时，一家竞争对手书商 Barnes & Noble 收购了 Borders 的部分资产，包括 Borders 的品牌商标和客户名单。然而，其第一个行动就是提醒客户他们的权利。我收到的 E-mail 是这样说的（部分内容）：

然而，重要的是你要明白，你有绝对的权利选择退出，不使你的客户数据转移到 Barnes & Noble。如果你选择退出，我们将确保从 Borders 收到的所有有关你的数据会以安全保密的方式处理掉。

这个例子说明了隐私政策是如何影响数据库管理员和企业数据专家的工作的。当然，也许你工作的公司永远不会破产，但你的公司可能会由于法律法规、经营状况或兼并等淘汰一些应用程序和数据。

最后，当应用程序完成它的使命时，DBA 要确定使用该数据的应用程序的最终状态。哪些数据不再需要了？其他的应用程序和进程也使用这些数据吗？是否有规则要求这些数据存储的时间比该应用程序长？<sup>②</sup>企业是否有任何明确的隐私政策强调了处理数据的特殊规则？

DBA 负责管理整体的数据库环境，通常包括安装 DBMS 和设置 IT 基础设施，以允许应

① <http://dataprivacyday2010.org/>。

② 法规对数据库管理的影响在第 15 章有深入的介绍。

用程序访问数据库。所有这些要在开发任何应用程序之前完成。除此之外，临时数据库访问也是许多企业的需求。

此外，DBA 还负责临时查询环境的建立，包括查询和报表工具的评估以及实施，建立工作准则和工作流程以确保临时查询的有效性，以及监控调优临时 SQL 语句。

如你所见，一名优秀的 DBA 和整个应用程序开发生命周期融为一体。DBA 数据方面的知识以及利用现代化的应用程序管理数据的方式都很受欢迎。

优秀的 DBA 和整个应用程序开发生命周期融为一体。

## DBA 生涯的一天

DBA 生涯的每一天都是非常忙碌的。他要做好多事情：不仅要维护生产环境和测试环境，同时还要关注现有的应用程序开发项目，参加策略和设计会议，帮助选择和评估新产品，把一些老旧系统联网。DBA Joe 刚才提交了“见鬼的查询又把系统搞崩溃了，你又能怎么样呢？”的评论。所有这些对于 DBA 来说都可能在同一天发生。

置身混沌之中，DBA 要知道所有事情，从技术到商业术语，再到最新的管理和技术的热门话题，这些他们都要知道。没有任何私人时间，DBA 在任何时候都要做好准备来回答任何问题，不仅是关于数据库的。

每当应用程序出现问题时，人们总是首先谴责数据库环境。数据库“有罪，且除了证明它无罪之外别无他法。DBA 从没有遇到过某个应用程序开发人员过来请教：“这里的 SQL 我用得很糟糕，你能帮忙解决吗？”一般是开发人员闯进 DBA 的工位吼道：“DB2（或者指责你最钟爱的 DBMS）有个问题，你为什么不把它解决掉，那样我很炫的程序就能跑起来了！”

应用程序出现问题时，数据库是“有罪的，直到证明是无辜的”。

因此，DBA 不得不去证明那个问题不是数据库引起的。他必须要知道信息技术的方方面面，在他所设计的 DBMS 和数据库架构中跟踪并排除错误。因此，他不仅要要是数据库技术方面的专家，还要知道和 DBMS 关联的所有 IT 元件：编程语言、操作系统、网络协议和网络产品、事务处理器、能想到的所有计算机硬件，以及更多。需要了解如此多样化的知识，DBA 也就成了非常宝贵的资源。他们的工作也变得有趣且充满挑战。

DBA 的工作是没有休息的。数据库程序是全天候运行的，所以 DBA 要时时做好处理问题的准备。DBA 大多随身携带呼机或手机，保证随传随到。如果数据库在凌晨 2 点出了问题，DBA 也要从床上爬起来清醒一下就去解决问题，让应用程序重新跑起来。否则会导致数据库故障，有可能导致业务流程完全终止。

DBA 经常周末都坐在电脑前，在数据库使用低谷时段维护并重组数据库。早上 9 点到下午 5 点期间维护数据库，不能让关键的数据库停止工作。坦白来讲，现在越来越多的企业都是 24 小时工作制，昔日的 8 小时工作制已经成为神话。

如果这对你还有吸引力，那么请继续往下读。实际上，也并不像听起来那么糟啦。这项工作很有趣，总是有新东西可以学，并且如前面提到的，薪水也不错。问题是，哪个人能 20 年甚至更多年一直做这种工作而不需要休息的？额，我好像听到你手机响了，你可能要去看有什么事了。

## 1.4 评估 DBA 的工作邀请

作为 DBA，在职业生涯中多次换工作几乎不可避免。每当要换新工作时，显然你都会考虑好多方面，如薪资、奖金、福利、调薪频率和休假时长。你还应该考虑那家公司是如何看待 DBA 的。不同企业对 DBA 的定位不同，寻找一家了解该工作的复杂性并一直学习该职位的需求的进取型企业对你的职业发展很重要。

下面是一些要问的有用的问题：

- 公司会为 DBA 提供常规的培训，让他们学习新 DBMS 的特征和功能吗？技术相关的培训怎么样？比如，编程、网络、电子商务、事务管理、信息队列等。
- 公司允许 DBA 定期参加本地用户组吗？<sup>①</sup>每年的远程用户组呢？
- 有备用 DBA，还是你会是一天 24 小时或一周 7 天待命的唯一一个？
- 是否有数据管理和系统管理组织，或者 DBA 也要履行所有这些职责？
- DBA 组会把应用程序开发团队看作合作伙伴吗？还是关系比较敌对？
- DBA 参与设计评审、预算讨论以及其他高级 IT 工作组和职务吗？

这些问题肯定的回答越多，这家公司 DBA 的工作环境就越先进。

## 1.5 数据库管理、数据管理和系统管理

一些企业分别为数据的商业方面和技术方面定义了不同的角色。数据的商业方面与数据管理是一致的，而更多技术方面都由数据库管理掌控。并不是每一家企业都有数据管理的职位，而许多企业都将数据管理并入数据库管理了。

许多企业都将数据管理并入了数据库管理。

有时企业也将数据管理的技术方面进行分离，DBA 负责使用 DBMS，而其他角色（系统管理或系统编程）负责安装并升级 DBMS。

### 1.5.1 数据管理

数据管理（Data Administration，DA）把数据资源管理的商业方面和用于数据管理的技术

---

① 即使它们发生在工作时间内？