Broadview
www.broadview.com.cn

韩国/日本/中国发行

海量数据库解决方案

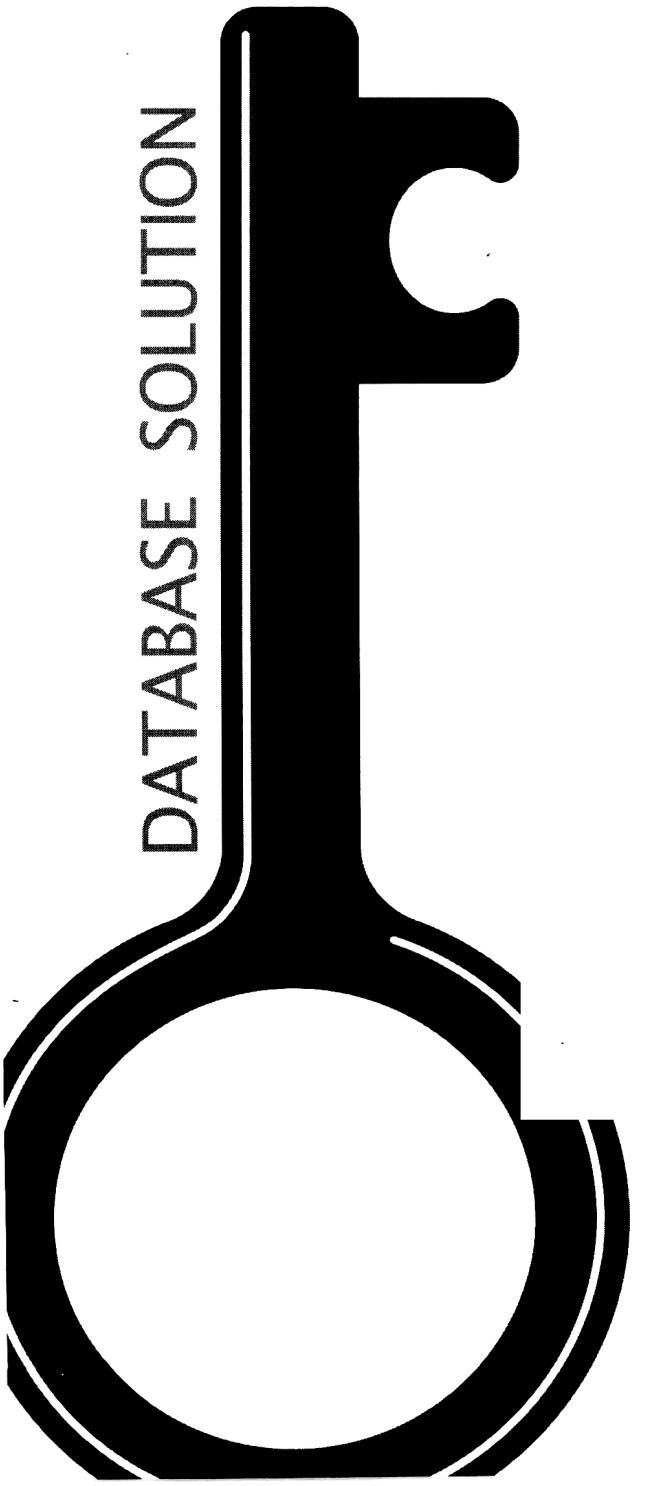
【韩】李华植 著
郑保卫 盖国强 译

涵盖数据库专家最新核心技术的RDBMS经典书籍

包含了将代码缩减为原来的1/10而速度提高至原来10倍的先进方法。

揭开了关系数据库的真面目，展示了截至现在未能被灵活使用的新技术。

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
HTTP://WWW.PHEI.COM.CN



海量数据库解决方案

【韩】李华植 著
郑保卫 盖国强 译
张乐奕 崔华 审校

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

本书将整体内容分为两部分，在第 1 部分中以影响数据读取效率的所有要素为类别，对其各自的概念、原理、特征、应用准则，以及表的结构特征、多样化的索引类型、优化器的内部作用、优化器为各种结果制定的执行计划予以详细说明，并以对优化器的正确理解为基础，提出对执行计划和执行速度产生最大影响的索引构建战略方案；在第 2 部分中主要介绍提高数据读取效率的具体战略方案，在这部分中介绍与数据读取效率相关的局部范围扫描的原理和具体应用方法，以及对被认为是提高数据库使用效率基础的表连接的所有类型予以详细说明。

《海量数据库解决方案》系列丛书深受广大读者的喜爱已经长达 10 年之久，在被誉为“圣经”的同时，它已经变成了数据库用户不可或缺的必读书籍。作者竭力探求能够让 IT 工作者在实际工作中轻松应用并掌控的巧妙方法，提供事半功倍的海量数据库解决之道。

本书适合数据库开发人员和数据库管理员等阅读。

版权贸易合同登记号　图字：01-2010-5909

图书在版编目（CIP）数据

海量数据库解决方案 /（韩）李华植著；郑保卫，盖国强译. —北京：电子工业出版社，2011.1
书名原文: New Written Large Scale Database Solution
ISBN 978-7-121-11883-8

Ⅰ. ①海… Ⅱ. ①李… ②郑… ③盖… Ⅲ. ①数据库系统 Ⅳ. ①TP311.13

中国版本图书馆 CIP 数据核字(2010)第 185498 号

责任编辑：李云静
印　　刷：北京东光印刷厂
装　　订：三河市鹏成印业有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编 100036
开　　本：787×1092　1/16　印张：28.75　字数：736 千字
印　　次：2011 年 1 月第 1 次印刷
印　　数：4000 册　定价：69.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

作者序言

这已经是第四次为本书写作者序言了，此时此刻过去 20 年的生活如同电影般在我的脑海里一一掠过。当我最初决定步入 IT 领域时就为自己立下了誓言，时至今日回想起多年走过的历程，其间充满了艰辛，也正是这无数的艰辛让我最终体验了收获的愉悦。

回望这 20 多年的足迹，我一直努力用新的视角去观察他人所忽视的领域，尝试用崭新的思维和充满创意的双手去耕耘。尽管如此，也仍然无法紧跟 IT 技术飞快的发展步伐。我为实现理想而终日不停前行的脚步，虽然忙碌但却无限满足。

众所周知，能够加工成宝石的原石比比皆是，一分耕耘，一分收获，每当我们初次接触某个新的东西时都会或多或少有些紧张。因此从这一层面来看，数据库散发着无穷的魅力，它如同渊博精深的智者般质朴，总是以真实、坦诚的心去面对每一位学习和研究它的人。

在过去并不短暂的岁月里我一直深信数据库的骨骼就是“数据”，并为这一理论的发展不断努力，吸收同仁们分享的经验而持续奋斗。为了打破始终在理论表面徘徊的固有模式而不断寻求新的尝试，并试图探求能够让 IT 工作者在实际工作中轻松应用并掌控的巧妙方法。

这种巧妙方法不能是只通过经验和试验才能获得的，它必须是利用日常常识就可以理解说明的方法。有这么一句话“会者不难，难者不会”，如果能够把一些复杂的理论与通俗浅显的常识相结合，那么不仅有利于人们的理解，更有利于人们在合适的情况下加以灵活运用。相反，有这么一句话“一知半解以为是”，意思是指那些只观其表不观其里就加以相信的人。

很多程序员只忠实地相信自己的经验，当问及为何如此时，大部分人的答案都是“因为我那样做过”或“那样比较好”。10 种类型的原理可以组合出 10 的阶乘（3 628 800）种现象，那么 100 种类型的原理所能够表现出来的现象数可以认为是一个天文数字。

如若仅凭经验去思考问题，无论怎么努力，最终也只能获得其中一部分的原理而已。然而，事实上我们是完全有能力深刻地理解这 100 种原理的。但如果不试图进行深刻钻研而只停留在表面，最终只能是一无所获。宝石是不会被轻易发现的，只有凭借最大的努力去寻找方能找到。

在不知不觉中当我们遇到了从表面上看无法解决的复杂问题时，会出现两种人：其一，是坚持不懈、彻夜不休也要寻找到最佳解决办法的人，这种人通过不懈的努力最终能够获得什么呢？事实上随着岁月的流逝，他们终将成为众人皆知的专家；其二，是认为过于烦琐，直接予以放弃的人，这种人只会让自己的血汗变成廉价的废弃物。

可以自豪地说“我付出了常人所无法想象的艰辛”，为了寻求完美的真理舍弃了很多

常人的生活。在没有钓到鱼时钓鱼人也许会为此而耿耿于怀，但在我看来问题的关键在于没有寻找到有效的钓鱼方法。如果钓鱼人能够充分理解我的想法，并甘愿为了改变自己的固有观念而付出较大努力，尽管他也可能会为此而花费大量的时间和心血，但坚信他一定能够获得别人所无法获得的成果。如果他研究出了别人所无法研究出的钓鱼方法，那么从此就再也不用为钓不到鱼而担心了。

各位读者在工作的同时究竟是否一直在使用一种平凡的方法呢？还是为了解决明天必须要完成的任务而临时抱佛脚呢？现在该到结束这种恶性循环的时候了。应用程序其实就是处理数据的手段而已，它需要紧跟流行的步伐，如不及时进行更新，在不经意之间就已经落伍了。

然而数据和数据库并非如此，不论岁月如何流失，我们积攒起来的“内功”是不会消失的。如果能对其原理有一个深刻的理解，那么不论何时何地都能够随心所欲地钓到很多鱼。随着数据库技术的发展进步，能够精确执行指令的 DBMS 与日俱增，随着对 DBMS 应用能力的不同所获得的性能差异使我们从技术中获得满足感。

不知不觉中《海量数据库解决方案》系列丛书深受广大读者的喜爱已经长达 10 年之久，在被誉为“圣经”的同时，它已经变成了数据库用户不可或缺的必读书籍。截至今日，本书依旧深受广大读者的喜爱。

IT 领域技术 10 年的发展状况与其他领域 100 年的发展状况相当，在数据库的发展历程中，有一些技术和原理被不断更新，有一些技术和原理被直接替代，也有一些技术和原理始终被维持使用。其中能够被持续使用而没有被改进的，说明它们是完美的，是适应性非常强的。

然而，我并没能及时将数据库所提供的新功能的真正含义和特性，及其在实际工作中的灵活运用方法和准则介绍给各位读者，借这次机会向各位忠实的读者表示深深的歉意。我知道很多读者对这本新书寄予了厚望， 相信通过我的努力能够让各位读者如愿以偿。

事实上，本人在此期间为了研究数据架构（Data Architecture）的相关理论花费了大量的时间和精力，因为我认为数据架构的重要性和根本性主要体现在它是搜集和管理数据体系的理论，它的目的在于以数据库技术为基础，构建更加深邃、全面的数据体系。现在，有很多用户开始对数据构建的相关理论感兴趣，并且一些用户还组织了专门的学习小组，我为此而感到非常欣慰。

随着信息化进程的不断加快，利用数据库所要管理的数据不仅会显著增多，而且也会变得非常复杂，由此而引发的数据合并、标准化、数据质量等方面的问题也已经到了不得不解决的境地了，实际上可以说是迫在眉睫。因此，构建以监督数据构架是否按照要求构建、完善对数据进行有效控制的元数据系统，已经成了我们所必须完成的迫切任务。

幸运的是，很多开发人员和管理人员都对数据构架注入了极大的关心，韩国政府为此专门设置了数据构架资格考试，也正是由于国家和各位 IT 领域从业者的努力才使得数据构架有了今天的良好发展形势。为了使其得到进一步的发展，我们为此而研究出了更

为体系化的理论方法，出版了高质量的书籍，设置了高标准的培训课程，构建出了将数据构架解决方案和相关组件相结合的元数据系统。

然而，在我为了推进数据构架的发展而付出大量时间和精力的同时，并没有放弃对数据库的研究。在此期间，我不仅连续不断地向很多用户提供数据库的咨询工作，而且也在不断地研究着新出现的技术。虽然我并没有公开地出版新书，但是却编写了大量的研究材料，并将其中相当大一部分在公司韩国数据库振兴院的主页上公开发表。尽管如此，现在对数据库的深刻研究，并将这些材料系统化的工作已经迫在眉睫了。

经过一年的昼夜工作终于完成了本书，并为了在本书中涵盖所有 DBMS 的“最小公倍数”而付出了很大的努力。虽然在各个 DBMS 之间多少都有一些差异，但是如果从原理和灵活运用准则的角度来看，它们之间其实并没有太大的差异。如若对所有 DBMS 的功能进行详细说明，反倒不利于各位读者的理解，所以在本书中以 Oracle 为基准进行说明的部分比较多。

尽管在本书中所使用的描述方法和命令语言主要以 Oracle 为基准，但所说明的原理和概念适用于所有的 DBMS，这就像尽管每一个照相机的具体操作方法都有所不同，但从拍摄照片的角度来看却并没有太大的差异一样。为了帮助各位读者更好地理解和应用本书中所介绍的原理和方法，介绍其被应用在各个不同 DBMS 中的相关书籍在不久的将来将呈现给各位读者。

本书将整体内容分为两个部分，在第 1 部分中以影响数据读取效率的所有要素为类别，对其各自的概念、原理、 特征、应用准则，以及表的结构特征、多样化的索引类型、优化器的内部作用、优化器为各种结果制定的执行计划予以详细说明，并以对优化器的正确理解为基础，提出了对执行计划和执行速度产生最大影响的索引构建战略方案。

在第 2 部分中主要介绍提高数据读取效率的具体战略方案。在这部分中，介绍与数据读取效率相关的局部范围扫描的原理和具体应用方法，以及对被认为是提高数据库使用效率基础的表连接的所有类型予以详细说明。

在第 1 部分的第 1 章中，主要对数据库在存储数据时所使用的基本存储结构进行详细说明。就像人类无法生活在水中，鱼类无法生活在陆地一样，结构上的特征对很多方面都有着根本性、决定性的影响。尽管不同的 DBMS 所使用的术语有所不同，但是当我们依据其本质进行分类时就会发现所有表现出来的形式都属于同一种类型。

在第 1 部分的第 2 章中，主要对优化器制定执行计划影响最大的所有索引类型进行详细说明。在此所涉及的索引类型有被广泛使用的最一般的 B-Tree 索引、在海量数据处理或数据仓库中能够获得较好效果的位图索引、基于虚拟列所创建的多样化的用户自定义函数索引等。

在第 1 部分的第 3 章中，主要对优化器进行深刻剖析。在此不仅介绍优化器在实现最优化操作时所执行的内部执行步骤，还对作为数据库处理数据步骤的执行计划的各种类型进行了分类说明。如果各位读者能够对各个具体的执行单位有一个很好的理解，那

么仍然能够对将其组合在一起的整体执行计划有一个很好的理解。另外，为了引导优化器按照我们所期望的方式制定执行计划，又对各种类型的提示（Hint）进行了说明。

在第 1 部分的第 4 章中，非常具体地提出了对制定最优化执行计划有着非常大影响的索引构建战略方案。优化器并不能开辟出新的读取路径，而只能在现有的读取路径中选择出比较有效的路径。因为按照本书中所提出的索引构建战略方案所构建出的索引能够改变现有的读取路径，所以如果没有依据此方案来构建战略性的索引，那么数据的读取效率必然会在很大程度上受到影响。

在第 2 部分的第 5 章中，以对执行计划的正确理解为基础，通过具体事例，对只读取了部分数据就能获得结果的秘诀——局部范围扫描的原理和多样化的应用类型进行了详细说明。在这里从一个新的高度对前面所介绍的方法进行了扩充说明。最后对在各种比较流行的留言板中实现局部范围扫描的方法进行了详细说明。

在第 2 部分的第 6 章中，对表连接的所有类型进行了详细说明。在这里对作为传统型的表连接方式——嵌套循环连接和排序合并连接，以及能够有效实现海量数据连接目的的哈希连接的相关原理进行了彻底剖析，并提出了多种灵活运用准则。另外，又对以多样化类型出现的半连接和在数据仓库中必须要使用的星型连接，以及星变形连接进行了详细说明。最后附加性地提出了位图连接索引的基本概念和灵活运用方法。

在这里向各位读者郑重承诺，我将在尽可能短的时间内完成其他两本系列书籍的编写工作。由于我经常在读者所期待的时间内未能出版约定的书籍，也许各位读者对我如期完成这两本书并不抱太大的希望，所以在这里希望各位读者能够谅解。我在管理自己企业的同时，既要研究新的技术，又要提供解决方案和编写书籍，时间的不足真的让我窘迫和无奈。

由于这两本系列书籍会同时在国外出版发行，所以不论我被多少琐事所困扰，都必须在约定的时间内完成编写工作。因此，这次我可以在这里向各位读者郑重承诺，在本书出版发行后我会以最快的速度完成这两本系列书籍的编写工作。

只凭借对方法的理解是无法征服数据世界的，即使将本书已经阅读了数十遍也只不过是停留在对其表层内容的理解上而已。所以希望各位读者能够对自己提出更高的要求，不要停留在对表层内容的理解上，而是以对本书中所介绍的所有原理的正确理解为前提，以本书中所提供的案例为参考，在适当的实际案例中加以灵活运用。我希望各位读者能够通过自己的努力和奋斗，发现别人所不能发现的新世界，而不是仅将自己局限于众人都能够看到的世界里。要勇敢地去挑战自己的极限，走向更大的成功。

2005 年 12 月 13 日

EN-CORE CONSULTING

代表咨询师 李华植

他山之石 可以攻玉

——《海量数据库解决方案》校订手记

编辑文案：

如果互联网也讲究人口红利，那中国无疑拥有得天独厚的人口基数与互联网普及速度。随着用户及服务规模的急速增长，海量数据库问题不期而至。然而，这一变迁过程进行得太快，相关从业人员来不及做好充分的技术准备，比如说，找不到任何一本可参考的书籍。

也正缘于此，曾服务过几乎所有本国一流世界级企业、拥有几十年从业背景的韩国数据库泰斗李华植先生的畅销著作《海量数据库解决方案》进入我们的视线。在充分了解到该著作在日韩经久不衰的事实，并有请国内知名数据库图书作家、技术权威盖国强老师谨慎评估后，我们有幸将其引入国内，供国内数据库同行参阅与品评。

国内韩版技术书籍的匮乏，主要是因为同时具备语言与技术能力的人选少之又少。为保证本书得以顺利出版，并保持较高水准，特邀原作者创办之 EN-CORE 公司的郑保卫先生进行初译，盖国强老师进行深入译校、中文统筹，张乐奕、崔华两位数据库资深专家与盖国强老师一道对本书进行了全面的技术解读、转译及校订。

本书的原著者、出版方及参与审译的同仁，都倾注了颇多心力，只希望本书的出版能在一定程度上填补国内关于海量数据库技术方面的空白，为日新月异的互联网产业发展贡献绵薄之力。实效如何尚不得而知，但请读者朋友不吝指正。

经过三个月的艰苦努力，我和张乐奕（Kamus）、崔华（Dbsnake）最终完成了《海量数据库解决方案》的校译工作。能和两位好友一起字斟句酌、逐字逐句地将韩国同人的名作校译引入中国，于我们于书都是一种缘分与机缘，在校订过程中恩墨科技的罗晓程也协助我们做了大量的文字审阅工作，在此致以深深的谢意。

在各位读者开始阅读本书之前，我们将接触、接受、校译这本书的感触与来龙去脉记录一下，供大家参详。

Eygle —— 推动技术交流与分享是一件功德

最早接触到这本书是在 2009 年，韩国 EN-CORE 公司的朋友找到我，希望我能够帮忙审阅并做些推荐，当时我正在新华社进行一个项目实施工作。

我当即表达了自己的几个观点：第一，如果是一本好书，我乐于阅读并做评荐，这是我的荣幸；第二，凡是促进技术交流与分享的事情，我都乐于支持并做出力所能及的

工作；第三，我愿意义务地去做这样一件事情。

韩国朋友非常愉快地表达了对我的谢意，并向我介绍了这本书：在韩国数据库界，该书的作者李华植是教父级的人物，而该书更是圣经般的读物，有累计几十万册的销量，读者不仅仅是数据库从业人员，各类技术人员都将其奉为经典和必读。

我相信了他们的说法，并且向韩国同仁表达了敬意。所有坚持不懈、执著于行的人，都值得我们尊敬。而如果能够促进亚洲数据库界的一些交流，那也是一件功德。

就这样我欣然答应了他们的请求，并期待看到这样一本书的出版。

当我拿到书稿时，一些新的问题出现了，我发现原译文太注重韩文的语言习惯，对于中国的读者来说，阅读起来会非常吃力，甚至会出现难于理解的局面。我建议他们做出进一步的修订，否则很难传达出作者的本意。基于双方的理解，他们诚挚地邀请我来完成这个审校工作，而经过慎重考虑之后，我认为一个人还不够，我需要一个团队，张乐奕和崔华成为了我的伙伴，这个团队通过了他们的考察，就这样，我们接下了这个比想象的还要艰辛的工作。

从原译文的传达，去理解作者的本意，这本就隔了一层，再加上韩文和中文的语言习惯不同，修订的工作极其艰苦，经常每个小时只能完成两三页的校对修正，然后我们三个人还要交叉校订。这期间的小插曲是，由于张乐奕是日语专业出身，他能够从该书的日文版借鉴不少东西，互相佐证。

就这样，三个月的时间转瞬即逝，我们的工作也已经接近尾声，我可以说的是，我还从来没有这么认真、字斟句酌地去读过一本书，而且是反反复复地阅读，这样阅读的一个好处是，我对这本书的理解和领会比任何其他一本书都要多。

书稿在手里就没有完美的一刻，我们还在不停地修正，直到出版前的最后一刻。我们真挚地希望能够有更多的读者喜欢它。虽然这期间我们已经尽了最大努力，但是仍然不知道读者会如何评价，所以我一直心怀忐忑。

关于这本书，可以说的是，我从中学到了很多东西，作者的很多理念让我受益匪浅，这些学到的东西将会指导和影响我以后的学习之路。

首先这本书并不是仅仅写给数据库从业人员的，作者期望所有对数据库感兴趣的读者都可以流畅地阅读、透彻地理解。所以，作者在本书中通过大量的类比、比喻将艰深的数据库知识与生活对应起来，使得平时很多不易理解的概念变得浅显易懂。基于深刻的积累，作者能够从不同角度对技术进行概括和阐释，往往有让人豁然开朗的别样感觉。比如在讲到局部范围扫描时，作者用排队等车用户的顺序以及出租车的出发时间进行类比譬喻，精到而浅显，任何人都可以一目了然地理解后面蕴含的复杂技术原理。这或许就是作者的本意，技术原本就是对完美生活的引申与抽象。

又比如在战略性索引构建一章，作者提到“只要为各个索引分配合理的任务，即使需要创建大量的索引也应当果断地作出决定”，为索引指定任务使得索引拟人化地变成了一个工人，有明确任务而不是平时我们茫然地创建索引。作者依据这样一个思想展开了

整章的内容，这也和我们的优化思路不谋而合，优化到极致的数据库，我们应当知道哪些 SQL 查询会使用到哪些索引，而索引又是为哪些查询服务的，必须详尽了解数据库与应用，才能对应用系统了如指掌、有的放矢。

读这本书，我们更多的是去理解作者高屋建瓴的数据库设计与优化思想，而不应该拘泥于具体的技术细节，比如作者所说的局部范围处理、战略性索引构建等，这些概念更多的是将细节的技术原理上升到生活道理与常规的思考。原本很多复杂的技术设计都是可以源自生活中浅显的道理，将技术、生活与常规的思考联系贯穿起来，在我看来，是作者带给我们的最大教益。

最初的一句承诺，导致了数月的锲而不舍，这期间的艰苦也让我们每个人都有所收获，当然更大的收获是我们几个朋友之间的友谊；感谢作者李华植带给我们的经典作品；感谢译者郑保卫，正是他的初始翻译才使得本书的中文版得以和广大读者见面；感谢博文视点的张春雨，他坚持不懈的沟通与推动促成了本书的最终出版。

愿这本书能为大家带来一些与众不同的感觉与收获！

盖国强

2010 年 7 月 22 日

Kamus——书读百遍其义自见

在今年（2010 年）一月份的时候，Eygle 跟我说有一本韩国 IT 届教父级的人物撰写的在韩国狂卖数十万册的数据库技术书籍，出版社希望引进到中国来。而这本韩文书已经有了初始中文译本，只是这份中文译本需要一些更熟悉 Oracle 数据库技术的人来进行再次修订。当第一次听到这件事情时，我以为那仅仅是校对，对于错别字的校正；对于技术术语把握的问题，我没有想到在真正开始着手以后会是如此的困难。

时间过了两个月，到了今年的 3 月份，正式拿到书稿，当开始修订我负责的第 3 章时，才发现这是一项巨大的工程。并不是说初始翻译不好，实际上已经很不错了，技术术语基本上也都准确，但是整段整段的句子读下来，却发现读一遍不明白在说什么，读两遍仍然似懂非懂，读三遍可能才知道，哦，应该是这个意思。而第 3 章是占全书比例最大的章节，有一百五十多页，而即使我全神贯注地去修订，进度也仅仅能够达到一个小时两到三页。之前我工作的环境是在客厅，前面是电视机，而为了修订这本书我不得不将工作环境转移到书房，因为哪怕是一点点声音的干扰也会让进度更加缓慢。这是我之前没有遇到过的情况，我一向自诩可以一心多用，但是修订这本书我遇到了极大挫折。究其原因，应该是初始翻译有不少是通篇大量同词语替换的情况，很多的修饰词都不符合中文的语言特点，而这些修饰词极大地干扰了我的阅读体验。另外一个重要的因素就是李华植先生在写作这本书的时候，对于很多他想极力阐述给大家的观点都辅以了大量的比喻，用出租车、用运动员、用下棋……而这些比喻的生动让阅读者会情不自禁地将

自己带入这个环境，去仔细琢磨这个比喻的后面想表达的真实含义。

为了能够更准确地理解李华植先生原文的意思，我甚至找来了本书的日文译本出版物（我大学本科学了四年的日语），在我实在不明白原中文译本的含义时，我会去对照日文是如何翻译这段的。有趣的是，通常这很有效果，我甚至根据日文译本在中文译本中添加了一些文字（也许是原中文译本遗漏了），以便于读者能够更好地理解文章含义。

很令人欣喜的是，在修订的后期我已经能够逐渐跟上李华植先生的思维，甚至对于他的比喻已经心有戚戚，修订的速度也得以提高。更令人欣喜的是，如此字斟句酌地阅读李华植先生的这本著作，让我在 Oracle 数据库性能优化的思想和策略上都学到了很多东西。“书读百遍其义自见”很能贴切地描述修订这本书的过程。

我、Eygle 和崔华在三个月中交叉修订，我修订完的第 3 章 Eygle 会再次修订一遍，然后传递给崔华，崔华再阅读一遍，同样 Eygle 修订完的第 4 章会传给我再重新阅读。我们如此努力，只是希望在最后这本书呈现在大家面前的时候，是一本符合中文阅读习惯的、不需要再像我们这样费劲就能够有所收获的数据库技术书籍。这是辛苦的三个月，但同样是很有意义的三个月。

对于每一位能够将自己的经验写下来分享给读者的作者，我都深深敬佩。我个人虽然一直没有这样的耐性坐下来去真正写一本书，但是我也一直在努力创造一种乐于分享、收获快乐的环境，创办 Oracle 中国用户组（www.acoug.org），每个月举办免费的活动，邀请演讲者来做技术分享，我一直在努力做自己力所能及的事情。

最后，即使我们如此修订这本书，一定还会有疏漏的地方，甚至是词不达意或者难以理解的部分，希望读者朋友们能够包容。但是更重要的一点是，在你觉得无法理解的时候，歇一段时间再去读第二遍、第三遍，要知道在我们修订的时候也是这样，“书读百遍其义自见”。

张乐奕（Kamus）

2010 年 7 月 4 日

Dbsnake——循序渐进始有成

三个多月的时间一晃就过去了，我们也终于完成了《海量数据库解决方案》的修订工作。

当初 Kamus 打电话找我，问我是否有兴趣跟他和 Eygle 一起去修订一本韩国书的时候，我几乎是不假思索地马上就答应了，当时其实我并不知道要修订的是什么书，也不知道具体的工作量会有多少。满口答应只是因为这是 Eygle 和 Kamus 找我——我没有理由拒绝。

修订这本书的时候恰逢我跟进的一个项目上线，白天项目的事情已经是忙得焦头烂

额了，修订的工作只能够放在晚上做。等到我真正开始修订的时候，才发现工作量真的是太大了，因为几乎是一直处于一种字斟句酌的状态，所以修订的进度非常缓慢，我常常是一个小时才能修订两到三页。怎么办？我怎样才能提高修订的速度，以至于不耽误整个团队的进度？没有别的办法，我只能压缩自己的睡眠时间，于是每个工作日我基本上都是 12 点左右入睡，早上 5 点半起床，就这样坚持了三个多月。

现在回头想想，我为什么能够坚持下来？除了要信守我自己的承诺之外，另外一个很重要的原因就是我对 Oracle 数据库太感兴趣了，再加上这本书确实是一本难得的好书！我在仔细修订这本书的过程中已经感觉到了自己的进步，这真的是一件令人非常兴奋的事情！

常常有朋友问我如何才能学好 Oracle？我这里想说的是我也不知道该如何才能学好，因为 Oracle 我也只懂一点点。但是当我看到李华植先生将其 20 年的数据库经验毫不保留地写在《海量数据库解决方案》中的时候，我觉得我还是应该把我的一点经验写出来，这样也许就能够帮助更多的人。

现在回想起来，我的 Oracle 入门过程大致分为四个阶段。

第一阶段：从 2004 年到 2006 年 8 月，我用了大概两年多的时间断断续续地看完了 OCP 9i 的一套培训教材（一共 4 本书）。很惭愧，用了两年多的时间才看完。但是大家要理解我，我本科和研究生都是学数学的，2004 年以前连最基本的 SQL 都不会写，而且那段时间项目的开发还是有一定的工作量的，我基本上只能够利用业余时间自学。

第二阶段：从 2006 年 8 月到 2008 年年初，因为那时候我的 Oracle 已经有了一点基础，所以我开始大量地看 Oracle 的各种官方文档，但是这些文档我大都只是过了一遍，唯一的例外就是《10gR2 Concepts》和《Oracle 10gR2 Clusterware and RAC Administration and Deployment Guide》，这两本书我都分别看过两遍。

第三阶段：从 2008 年年初到现在，我开始在 metalink 上大量阅读文章，在我真正开始接触 metalink 的时候我才发现，哇塞，这是多么精彩的一个世界！在我看来，metalink 就是最好的老师了！直到现在，我工作的第一件事情依然是把 metalink 打开，平常没事就泡在上面，迄今为止，我已经在 metalink 上看过超过 1500 篇文章了。

第四阶段：从 2009 年 3 月到现在，我开始仔细阅读 DSI（Data Server Internals）教材。DSI 其实我看的并不多，迄今为止，只看过了如下几本：

DSI303-Advanced Backup,Restore and Recovery Techniques

DSI401-Dumps Crashes and Corruptions

DSI402-Space and Transaction Management

DSI402e-Data types and block structures

DSI403e-Recovery Architecture Components

DSI404e-query_optimizer

DSI405-Performance Tuning

总的感觉是 DSI 没什么，真的没有我想象的那么难。但是这里我想说的是 DSI 应该在你具有一定的基础后才开始看，早看反而无益！

前前后后六年的时间，当我看完 DSI405 后，我真的觉得自己的 Oracle 已经入门了。其实细心的朋友已经可以发现，我的 Oracle 入门过程用一句话来概括就是“循序渐进”。

当我看完 DSI 系列并且在我的个人网站（www.dbsnake.com）上撰写了 100 多篇文章后，终于迎来了属于我自己的机会——我结识了 Eygle 和 Kamus，并且有机会和他们一起开始迎接许多梦寐以求的挑战。

一直以来，Eygle 和 Kamus 都是我努力的目标，修订的这三个多月虽然辛苦，但是能和自己的知音益友一起做事情，于我而言是非常荣幸也是非常开心的事情，再辛苦又算得了什么。这次我修订了第 1 部分的第 1 章、第 2 章和第 2 部分的第 2 章，交叉修订了第 1 部分的第 3 章。在修订过程中，我付出了大量的努力，以自己所学、所知去理解并传达作者的真知灼见。然而再怎么尽心竭力，我们的工作也难免存在种种不足，但是我们确确实实已经尽力，剩下的只能交给读者去评判吧。

虽然这次修订的过程是极其痛苦的，在殚精竭虑之后常常会感觉很崩溃；但是如果能有机会让我继续修订《海量数据库解决方案》的后续书籍，我非常愿意再多崩溃几次。

崔华（Dbsnake）

2010 年 6 月 29 日

好了，这就是几位校译者的心声。现在这本书已经翻越了山川、民族与语言的障碍，来到中国读者的面前，让我们一起来阅读和领会一下来自亚洲数据库界的声音吧！

目 录

第 1 部分 影响数据读取的因素