

Delivering Business
Intelligence with Microsoft
SQL Server 2005

Microsoft
SQL Server 2005
商业智能实现

数据库专业人员的必备技能

- 收集、转换与分析不同类型的企业数据
- 学习商业智能的完整生命周期
- 实现应用的商业智能

(美) Brian Larson 著
赵志恒 武海锋 译
杨大川 审校

Microsoft SQL Server 2005

商业智能实现

(美) Brian Larson 著

赵志恒 武海锋 译

杨大川 审校

清华大学出版社

北 京

Brian Larson

Delivering Business Intelligence with Microsoft SQL Server 2005

EISBN: 0-07-226090-4

Copyright © 2006 by The McGraw-Hill Companies, Inc.

Original language published by The McGraw-Hill Companies, Inc. All rights reserved. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Simplified Chinese translation edition is published and distributed exclusively by Tsinghua University Press under the authorization by McGraw-Hill Education(Asia) Co., within the territory of the People's Republic of China only (excluding Hong Kong, Macao SAR and Taiwan). Unauthorized export of this edition is a violation of the Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书中文简体字翻译版由美国麦格劳-希尔教育出版(亚洲)公司授权清华大学出版社在中华人民共和国境内(不包括中国香港、澳门特别行政区和中国台湾地区)独家出版发行。未经许可之出口视为违反著作权法,将受法律之制裁。未经出版者预先书面许可,不得以任何方式复制或抄袭本书的任何部分。

北京市版权局著作权合同登记号 图字: 01-2006-6736

本书封面贴有 McGraw-Hill 公司防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话: 010-62782989 13701121933

图书在版编目(CIP)数据

Microsoft SQL Server 2005 商业智能实现/(美)拉森(Larson,B.)著;赵志恒,武海锋 译;杨大川审校.
—北京:清华大学出版社,2008.3

书名原文: Delivering Business Intelligence with Microsoft SQL Server 2005

ISBN 978-7-302-16922-2

I. M… II. ①拉…②赵… III. 关系数据库—数据库管理系统, Microsoft SQL Server 2005
IV. TP311.138

中国版本图书馆 CIP 数据核字(2008)第 009201 号

责任编辑: 王 军 李 阳

装帧设计: 孔祥丰

责任校对: 成凤进

责任印制: 王秀菊

出版发行: 清华大学出版社

<http://www.tup.com.cn>

c-service@tup.tsinghua.edu.cn

社 总 机: 010-62770175

投稿咨询: 010-62772015

地 址: 北京清华大学学研大厦 A 座

邮 编: 100084

邮购热线: 010-62786544

客户服务: 010-62776969

印 装 者: 清华大学印刷厂

经 销: 全国新华书店

开 本: 185×260 印 张: 39.75 字 数: 824 千字

版 次: 2008 年 3 月第 1 版 印 次: 2008 年 3 月第 1 次印刷

印 数: 1~4000

定 价: 79.80 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话: (010)62770177 转 3103 产品编号: 022759-01

译者序

随着经济的不断发展，企业所要面临的挑战越来越多，企业需要在竞争中保持主动、增加盈利。为此，需要准确了解公司的运营状况，在此基础上作出合理的预测、制定具体的行动计划。今天，广泛渗透并不断发展的信息技术，使得我们不仅可以高效地维护各类业务数据，实现管理自动化，还可以集合多个部门，以及不同类型的企业数据，实现快捷的综合业务分析，为管理决策提供准确有效的信息。商业智能的兴起正反映了软件领域对这一市场需求的响应。Microsoft 公司研究了不同企业的多种需求，经过五年的开发，为我们奉献了一套全新的商业智能软件——SQL Server 2005。SQL Server 2005 可以快速构建各类商业智能解决方案，提供深入的业分析统计和监控预测平台，进一步推动企业的信息化管理和业务发展。

本书作者 Brian Larson 是 Microsoft 认证的解决方案专家(MCSD)和数据库管理员(MCADB)，也是 Reporting Services 开发小组的成员，具有丰富的数据库系统设计和商业智能开发经验。本书是作者对 Microsoft SQL Server 2005 研究的最新体现，将带您领略 Microsoft 新一代商业智能平台的体系结构。

作者根据技术学习的一般规律，通过精心设计的学习曲线，从商业智能的基础知识、商业智能的数据来源及作用入手，循序渐进地介绍了统一维度模型以及商业智能开发的基本理论、技术体系及步骤。全书以 Maximum Miniatures 公司为示例，以精心设计的示例进行阐述，几乎每一个知识点都结合实例从应用的角度进行了详解。通过对示例的学习配合相应的实践步骤，一步一步地向读者展示了如何创建高效实用的 BI 解决方案。

本书由赵志恒、武海锋翻译。在翻译过程中，我们对原书中存在的一些明显错误进行了修改，对于少数不太清楚的地方也参考了一些相关资料。尽管我们作了许多努力，但由于水平有限，在技术上仍有理解不够深入的地方。书中存在的疏忽与不当之处，恳请读者批评指正，请将批评和建议发送到电子邮箱 wkservice@tup.tsinghua.edu.cn，我们将不胜感激！

译者
2007 年 7 月

审校者序

本书的英文名为“Delivering Business Intelligence with Microsoft SQL Server 2005”。如果直译的话,“Delivering”可以翻译为传递、释放、展现,等等。那么,到底应该如何理解呢?

首先,我们要了解什么是商业智能(Business Intelligence)。

有人说,商业智能是一个软件。有人说,商业智能是一种挖掘技术。有人说,商业智能就是决策支持系统(Decision Support System)。也有人说,商业智能就是数据挖掘。

实际上,商业智能是一种解决方案,它的目的是把用户积累下来的大量数据转化为业务人员容易理解的信息,进而辅助决策。

数据库中存在的数据(Data),对于业务人员来说,只是一些无法看懂的天书。他们真正需要的是信息(Information)。那么,我们如何解决这个矛盾的呢?简单的方案是报表系统;深一个层次的商业智能是多维数据分析(OLAP);更深一个层次的商业智能叫作数据挖掘(Data Mining)。因此,很难用一个单独的产品或者技术来概括商业智能,我们只能说,商业智能是一条路,一条从数据到信息的路。

BI = Data → Information

显然,随着现代社会中数据的爆炸式增长,商业智能的价值也越来越得到了体现。过去,在从数据到信息之路的每一个环节上,很少有一个相对完整、统一的技术方案。我们往往需要寻找不同公司的不同产品,然后进行复杂的系统集成。自从 Microsoft SQL Server 2005 发布之后,对于商业智能工程师来说,无疑是一个好消息。因为微软的这一款产品(严格说是一组,多个产品),实际上囊括了从数据整合,数据清洗,数据仓库与集市,一直到报表,多维数据库,数据挖掘等每一个环节。

因此,“Microsoft SQL Server 2005 商业智能实现”就成了本书最终的中文译名,也恰恰代表了我们对读者的祝愿:希望对本书的学习能够给您带来商业智能蕴含的巨大价值。

与其他商业智能书籍相比,本书的一个特点就在于使用了大量详尽的示例,几乎每一个知识点都有精心安排的具体步骤。因此,非常适合于技术人士自学。

审校过程中,我们看到 Excel 2007 已经发布,并且在商业智能方面有较强大的功能,因此,我们把本书中的相应章节更新到了 Excel 2007 版本,以确保技术的及

时、先进。

我从 2003 开始创建的北京迈思奇科技有限公司致力于将国外的先进商业智能技术引进国内，帮助国内的企业实现商业智能的价值。公司成立四年来，与微软密切合作，通过近百次讲座和培训，为企业培养了大量的商业智能专业人员；同时，在承担商业智能项目开发的过程中，公司也积累了优秀的团队和丰富的项目案例，创立了自己的品牌。

感谢清华大学出版社，及时为我们引进了这部优秀的教材；感谢本书的译者，准确而清晰地传达了原著的精华；也感谢迈思奇公司参与本书审校工作的各位 BI 咨询顾问。

北京迈思奇科技有限公司 杨大川
2007 年 10 月

目 录

第 I 部分 商业智能化

第 1 章 使组织机构具备有效决策能力	3
1.1 有效决策能力	4
1.1.1 谁是决策者?	4
1.1.2 什么是有效决策?	5
1.2 具备有效决策能力的要素	5
1.2.1 量化的目标	6
1.2.2 具体的度量	7
1.2.3 适时的信息	8
1.3 商业智能	10
第 2 章 充分利用现有资源	13
2.1 商业智能的作用	14
2.1.1 用于问题明确的情况	14
2.1.2 发现新问题并解决问题	15
2.2 多层次的商业智能	15
2.2.1 金字塔的顶层	16
2.2.2 中间层次	17
2.2.3 最低层次	18
2.3 Maximum Miniatures 公司	18
2.3.1 业务需求	19
2.3.2 当前系统	19
2.4 建立基础	21
第 3 章 商业智能的源系统	23
3.1 寻找源系统	24
3.2 数据集市	26
3.2.1 数据集市的特征	27
3.2.2 数据集市的结构	29
3.3 小结	36
第 4 章 统一维度模型	39
4.1 联机分析处理	40
4.1.1 使用多维数据集建立 OLAP 系统	40
4.1.2 OLAP 系统的特性	43
4.1.3 OLAP 系统的体系结构	45
4.1.4 缺点	46
4.2 统一维度模型	47
4.2.1 结构	48
4.2.2 优点	52
4.3 小结	53
第 5 章 开始商业智能的第一步	55
5.1 Business Intelligence Development Studio	56
5.1.1 Visual Studio	56
5.1.2 Business Intelligence Development Studio 导航	57
5.1.3 Business Intelligence Development Studio 的选项	70
5.2 SQL Server Management Studio	73
5.3 小结	81

第 II 部分 定义商业智能的结构

第 6 章 建立基础——创建和填充数据集市	85
6.1 数据集市	86
6.2 设计数据集市	88
6.2.1 决策者的需求	89
6.2.2 可用数据	90
6.2.3 数据集市的结构	90
6.2.4 使用 SQL Server Management Studio 创建数据集市	99
6.2.5 使用 Business Intelligence Development Studio 创建数据集市	112
6.3 小结	120
第 7 章 使用 Integration Services 填充数据集市	121
7.1 Integration Services	122
7.1.1 包结构	122
7.1.2 包项	139
7.2 包的开发特性	200
7.2.1 开发 Integration Services 包	200
7.2.2 包开发工具	208
7.2.3 迁移 SQL Server 2000 中的 DTS 包	217
7.3 应用 Integration Services 包	218
7.3.1 部署 Integration Services 包	218
7.3.2 执行 Integration Services 包	220

7.4 小结	221
--------	-----

第 III 部分 分析多维数据集

第 8 章 多维数据集的机制——度量和维度	225
8.1 回顾与展望	226
8.1.1 完善生产线数据集市	226
8.1.2 创建多维数据集	236
8.2 度量	242
8.2.1 度量组	243
8.2.2 计算成员	245
8.2.3 度量聚合	248
8.3 维度	252
8.3.1 管理维度	252
8.3.2 建立维度和度量组的联系	253
8.3.3 维度类型	256
8.3.4 渐变维度	259
8.4 小结	263
第 9 章 OLAP 多维数据集的属性	265
9.1 准备部署多维数据集	266
9.1.1 部署和处理	266
9.1.2 从 Business Intelligence Development Studio 进行部署	267
9.1.3 从“Analysis Services 部署向导”进行部署	273
9.2 其他的多维数据集属性	279
9.2.1 链接对象	279
9.2.2 商业智能向导	282
9.2.3 关键性能指标	283
9.2.4 操作	288
9.2.5 分区	290
9.2.6 透视	306

9.2.7 翻译	307
9.3 小结	308
第 10 章 编写新的脚本——MDX	309
10.1 术语与概念	310
10.1.1 我们在哪里?	310
10.1.2 从当前位置到达另 一个目的地	326
10.2 运用 MDX 脚本	333
10.2.1 多维数据集的安全	333
10.2.2 今年和去年的对比 以及年初到本日止 的汇总	343
10.3 从多维数据集中 提取数据	347
第 11 章 MDX 查询	349
11.1 MDX SELECT 语句	350
11.1.1 基本的 MDX SELECT 语句	351
11.1.2 其他查询工具	362
11.1.3 其他维度	372
11.2 其他查询语法	374
11.2.1 运算符	374
11.2.2 函数	377
11.3 数据挖掘	380

第IV部分 数据挖掘

第 12 章 数据挖掘入门	383
12.1 什么是数据挖掘	384
12.1.1 从混乱到有序	384
12.1.2 数据挖掘完成 的任务	387
12.1.3 数据挖掘的步骤	392
12.2 数据挖掘算法	394

12.2.1 Microsoft Decision Trees 算法	394
12.2.2 Microsoft Naive Bayes 算法	395
12.2.3 Microsoft Clustering 算法	396
12.2.4 Microsoft Association 算法	397
12.2.5 Microsoft Sequence Clustering 算法	399
12.2.6 Microsoft Time Series 算法	400
12.2.7 Microsoft Neural Network 算法	402
12.3 小结	403
第 13 章 处理数据挖掘模型	405
13.1 数据挖掘结构	406
13.1.1 数据列	406
13.1.2 数据挖掘模型	406
13.1.3 训练数据集	407
13.2 挖掘模型查看器	418
13.2.1 Microsoft 决策树	418
13.2.2 Microsoft Naive Bayes	423
13.2.3 Microsoft 聚类	426
13.2.4 Microsoft 神经网络	429
13.2.5 Microsoft 关联	430
13.2.6 Microsoft 序列聚类	432
13.2.7 Microsoft 时间序列	432
13.3 小结	433
第 14 章 使用数据挖掘	435
14.1 “挖掘准确性图表” 选项卡	436
14.1.1 “列映射”选项卡	436
14.1.2 提升图	441

14.1.3	利润图	445
14.1.4	分类矩阵	446
14.2	“挖掘模型预测”	
	选项卡	447
14.2.1	单独查询	447
14.2.2	Prediction Join	
	查询	451
14.3	Data Mining Extension	456
14.3.1	预测查询语法	457
14.3.2	预测查询的类型	458
14.4	数据传递	463

第 V 部分 报表

第 15 章	使用 Reporting Services	
	实现商业智能	467
15.1	Reporting Services	468
15.1.1	报表结构	469
15.1.2	传递报表	470
15.2	报表服务结构	471
15.2.1	Report Server	471
15.2.2	Reporting Services	
	的组成部分	473
15.2.3	Reporting Services	
	的安装注意事项	476
15.3	创建报表	478
15.4	Report Manager	526
15.4.1	文件夹	526
15.4.2	Report Manager	527
15.4.3	使用 Report Designer	
	部署脚本	528
15.4.4	使用 Report Manager	
	上传报表	529
15.4.5	在 Report Manager	
	中打印报表	534

15.5	在 Report Server 上	
	管理报表	536
15.5.1	安全性	536
15.5.2	链接报表	543
15.5.3	报表缓冲	544
15.5.4	执行快照	547
15.5.5	报表历史记录	548
15.5.6	标准订阅	548
15.5.7	数据驱动订阅	549
15.6	即席报表	550
15.6.1	报表模型	551
15.6.2	Report Builder 基础	558
15.7	小结	562

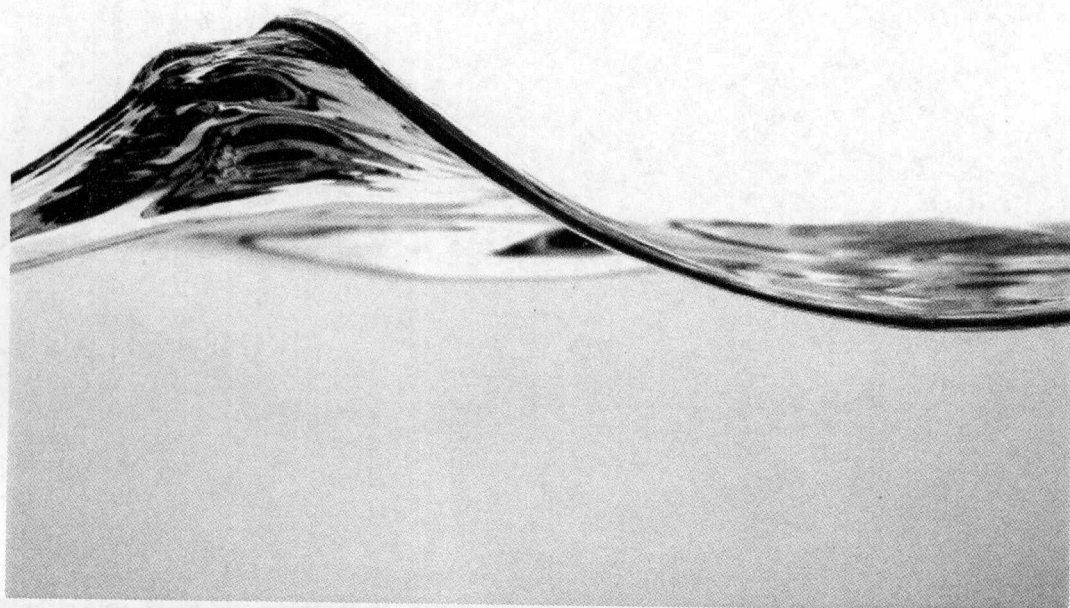
第 16 章	在应用中整合 OLAP	563
16.1	ADOMD.NET	564
16.1.1	ADOMD.NET	
	的结构	564
16.1.2	ADOMD.NET 示例	566
16.2	不带 Report Manager 的	
	Reporting Services	572
16.2.1	URL 访问	572
16.2.2	Web 服务访问	587
16.2.3	Report Viewer 控件	592
16.3	小结	600

第 17 章	Excel 透视表和透视图表	601
17.1	Excel	602
17.1.1	创建透视表和	
	透视图表	602
17.1.2	透视表	603
17.1.3	透视图表	612
17.1.4	其他功能	621
17.2	小结	621

附录 A	学习资源	623
------	------	-----

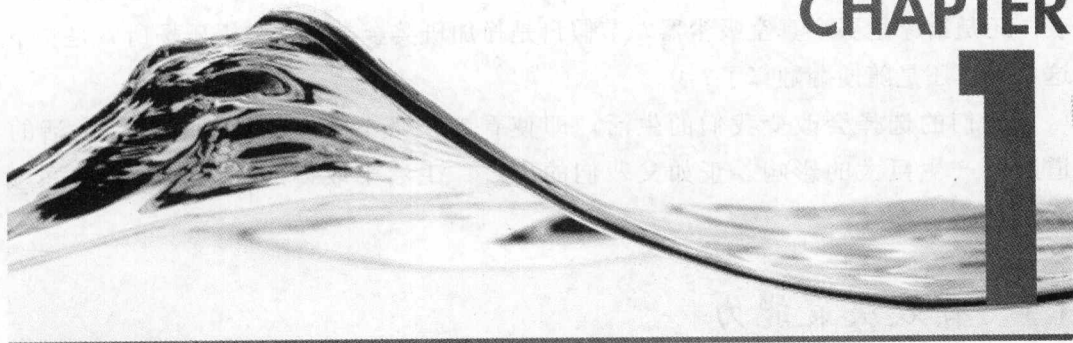
PART

商业智能化



外語專業





CHAPTER

1

使组织机构具备有效决策能力

本章内容：

- 有效决策能力
- 具备有效决策能力的要素
- 商业智能化

人生充满了选择。午餐吃油炸三明治还是吃沙拉？是给自己买块期望已久的手表，还是给家里买个真空吸尘器？节假日是挣加班奖金，还是去娱乐场所？是买下这本书，还是随便翻翻算了？

我们的选择会改变我们的生活。即使看起来微不足道的选择也会在生活的道路上产生巨大的影响。正如父辈们的忠告：生活中取得成功的关键是做好每个决定。

1.1 有效决策能力

良好的决策能力对于我们的工作与人生同样重要。每天我们所工作的组织机构都要做出很多影响方向和效率的决策，它们涉及到生产、市场以及人事等诸多事情。每个决策都会影响到成本、销售以及利润。正如我们的个人生活一样，组织机构成功的关键也在于良好的决策。组织机构必须具备有效决策能力。

1.1.1 谁是决策者？

在我们的组织机构中，谁负责进行决策？直观印象中好像应该是 CEO、主管、主席等那些必须担当有效决策的机构最高层人士，如果那些人能够做出适当的战略决策，这个机构就能走向成功。

然而事情并非如此简单。历史上无数的例子说明，往往由于执行过程中许多糟糕的决策而导致辉煌的战略计划破灭。拿破仑在比利时的战斗中曾经也有相当好的战略计划，然而由于他的元帅们错误决策，最终导致了滑铁卢惨败。

这么看来，也许同样重要的是让所有上层的管理者都成为有效决策者。包括 CFO、CIO、副主席、一般委员以及部门负责人都要在制定方针以及确定战略计划执行优先权时，做出良好的决策。如果所有的上层管理机构的领导都能做出有效的决策，那么机构应该就有了成功的保证。

事实上也不一定如此。上层给出的有效计划与方针在付诸实施的时候，也可能由于传达的错误而没有执行决策。所以，应该让机构的上上下下都拥有有效的决策。如果下层管理者的努力不会被高层领导否决，他们就会有更高的士气并能投入更多的精力。此外，公司高层的成功在很大程度上不仅是有效决策的成功同时也是底层人员的成功。只有把所有层次的有效决策结合起来才能通向成功之路。

1.1.2 什么是有效决策?

如果有市场需要的产品或者服务,在正确的时间、正确的地方,产生适当的开销,同时有必要的客户支持,这样的机构就能够成功。很明显,任何商业计划或者任务陈述大概都是阐述如何保证正确决策的。

所不明确的是,机构怎么才能保证提供良好的产品或服务。答案是让每个层次都有人来负责有效决策。那么什么是有效决策呢?



定义

有效决策是能够适时地保证组织机构朝着一系列既定目标前进的决定。

有效决策可以适时地保证机构向目标前进。这个定义有些泛泛而谈。事实上,这通常只能是一个不错的口号,因为它太广泛了,不能在日常操作中起多大作用。使用这个定义,我们可以定义能够进行有效决策的三个要素。

- 首先,必须要有一组前进的目标。
- 其次,必须要有一个方法以衡量选定的路线是走向或是背离这些目标。
- 再有,必须将基于这些衡量方法的反馈信息适时地提供给决策者。

这些信息同时作为初始决策的基础信息以及决策结果的反馈信息。定义有效决策能力虽然很容易,但要应用这个相当模糊的定义并转化为具体的业务实践却还需要做很多工作。



定义

基础信息是在决策过程中做出正确决策的基础。



定义

反馈信息用来在决策制定之后评估决策的有效性。

1.2 具备有效决策能力的要素

在前一节中,我们知道了用于有效决策的有三个要素:特定的目标,具体的度量,以及适时的基础和反馈信息,如图 1-1 中所示。本节将详细讨论每个要素,研究如何促进有效决策能力。

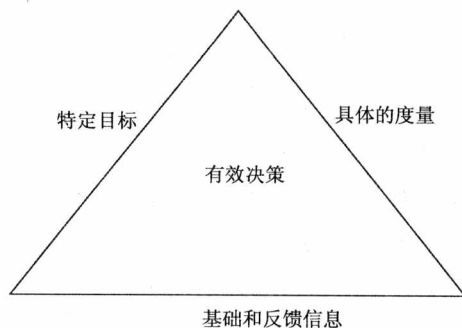


图 1-1 有效决策能力的三个要素

1.2.1 量化的目标

在 Mel Brook 的电影 *the producers* 中, Max 和 Leopold 上演了一个开始时就注定要失败的百老汇音乐剧, 借此使他们能逃避投资者的债务。除了这种特例外, 机构一般不会去寻求失败。人们聚到一起、筹款、建立机构体系、制定商业计划、进行无数次会议, 所有的这些都是为了成功。对于很多机构而言, 最重要也最致命的问题是, 他们没有定义到底什么是成功。他们不知道目标是什么。

通常, 机构在任务陈述中都有很多模糊的目标。类似“超级客户满意”、“增加净利润”或者“更好地符合社会需求”等措辞的确让招待会、年度报告以及股东会议报告增色很多。但这些仅仅是适用于市场中寻求合作关系或者雇员之间产生集体精神的口号。它们不能产生衡量事业成就的实质内容。

“超级客户满意”是一个很完美的目标(即便仅有一半的公司表示了要向“超级客户满意”的目标前进, 并确实进行了努力, 那么这个世界也会更加快乐)。问题是如何来衡量“客户满意”。我们怎样才能知道已经达到了这个目标或者正在朝这个方向前进? 我们需要一些更实质性的内容以及更多可度量的参数。

除去漏洞百出的“超级客户满意”, 更好的目标可能是“维持超级客户满意度, 由重复的客户订单来衡量, 要达到 80% 重复订单的目标”。这个目标可能还需要更具体一些, 但是这是可度量目标的开端。基于重复订单的增长或者下降, 我们可以衡量决策是否正在推动机构朝正确的方向前进。

“增加净利润”可以让股东满意。但是机构仍需要确定哪些运作成本影响了净利润, 同时在众多协作方案中如何分配运作成本。我们还要说明回报投资者的净利润增长率。1% 的增长已经会让股东脸上露出微笑, 5% 的目标是否会更好些? 一旦将这些细节添加进去, 我们就有了一个前进的具体目标。

“更好地符合社会需求”是一个高尚的目标。但是这个社会需求什么，同时如何知道在什么时候符合了社会的需求？实际上，我们需要选择一项具体化的社会需求度量，比如增加普通收入家庭可承担的高品质住宅的数目。然后我们可以定义什么是普通收入家庭可承担的高品质住宅以及需要达到多少的增长目标。

作为有效决策能力的一部分，一个目标必须

- 包括一个特定的目的；
- 提供一种方法来衡量我们是否正在朝这个目的前进。

通过图 1-2，我们知道，目标要有一个用于瞄准的靶心和一个对目标接近程度的评分方法。

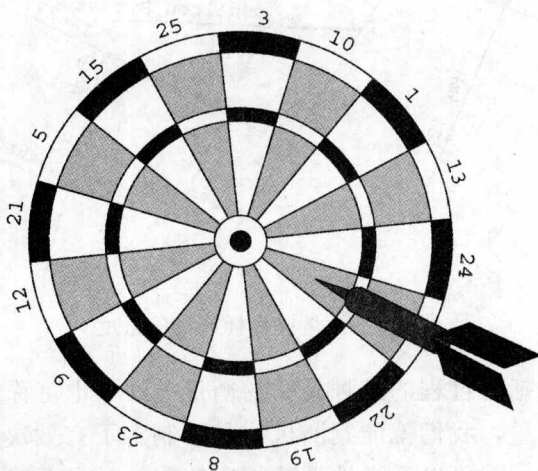


图 1-2 有效决策能力中目标的组成元素

1.2.2 具体的度量

目标是重要的。实际上正如前面讨论的，它们是有有效决策的基础。然而如果不朝着目标前进，目标就毫无用处。只有当人们开动引擎奔向目标时，才有可能达到终点。

接下来就要展现决策能力。每个决策都会推动公司朝某个方向前进。一些决策使公司产生了巨大的改变。这些通常是高层管理部门制定的方针和优先级决策。这些决策制定了机构将要在一个较长时间段、一个财务年度、一个学期、或者机构存在的整个时间段中采用的一般路线。基本上它们都是引导机构通向目标的决策。

而也有一些决策只是使机构的某些小的方面有相对较小的发展。这些决策包括工作组优先权决策和日常运作决策，甚至可能小到某个雇员处理某个客户投诉的决策，或者某个销售代表对下一个要拨的电话号码的决定。这些机构方向上较小的变