

信息科学与技术丛书

王 飞 刘国峰 编著

商业智能深入浅出

第2版

大数据时代下的 架构规划与案例

- 商业智能概述
- 商业智能数据仓库系统应用实例
- 数据仓库理论知识
- 大数据架构与实践
- 数据挖掘和分析
- 电力行业和金融行业的商业智能案例



机械工业出版社
CHINA MACHINE PRESS

信息科学与技术丛书

商业智能深入浅出

——大数据时代下的架构规划与案例

第2版

王 飞 刘国峰 编著



机械工业出版社

本书虽然是《商业智能深入浅出——Cognos, Informatica 技术与应用》一书的修订版,但在结合各方面的反馈意见之后,对内容上做了很多调整,力求最新、最细。同时书中将商业智能和大数据有机地结合起来,增加了一些相应的章节和案例,扩充了知识点,几乎涵盖了商业智能领域的所有知识。本书并不推崇细节性的理论知识讲述,因为每一部分理论都可以写成一本书。本书主要针对如何解决项目中所遇到的问题,以及商业智能项目开发的一般流程。本书还力图帮助初学者快速进入到项目之中,所以本书对他们来说具有极高的参考价值。

本书内容可以分成如下几个部分。

理论篇:主要包括商业智能概述、数据仓库理论知识、数据挖掘和分析、大数据理论知识等内容。

项目篇:主要包括商业智能项目需求的定义、商业智能项目模型的建立、商业智能数据仓库系统应用实例、大数据架构与实践、电力行业和金融行业的商业智能案例等内容。

工具篇:详细介绍两大商业智能工具 Cognos 与 Informatica 的理论知识和使用方法。

实践篇:包括 Cognos 报表的开发、部署和实践等内容。

本书还附赠了部分源代码和一些有价值的文档模板。

图书在版编目(CIP)数据

商业智能深入浅出:大数据时代下的架构规划与案例/王飞,刘国峰编著.

—2版. —北京:机械工业出版社,2014.8

(信息科学与技术丛书)

ISBN 978-7-111-47601-6

I. ①商… II. ①王… ②刘… III. ①企业管理—应用软件 IV. ①F270.7

中国版本图书馆CIP数据核字(2014)第181018号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

策划编辑:丁 诚

责任编辑:丁 诚 周 萌 责任校对:张艳霞

责任印制:李 洋

三河市宏达印刷有限公司印刷

2014年9月·第2版第1次印刷

184mm×260mm·30印张·744千字

0001—3500册

标准书号:ISBN 978-7-111-47601-6

定价:79.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

电话服务

网络服务

社服务中心:(010) 88361066

教材网:<http://www.cmpedu.com>

销售一部:(010) 68326294

机工官网:<http://www.cmpbook.com>

销售二部:(010) 88379649

机工官博:<http://weibo.com/cmp1952>

读者购书热线:(010) 88379203

封面无防伪标均为盗版

前 言

谁需要阅读本书

简而言之，如果您是一直关注商业智能和大数据领域的开发人员、架构师、咨询顾问或者是相关专业的师生，无论是初学还是已有一定工作经验，这本书都适合您。

考虑到很多读者刚刚涉足这个领域，所以本书不会使用晦涩难懂的语言，并尽量用浅显易懂的方式去说明问题。在内容方面，本书几乎涵盖了商业智能的方方面面，包括一些架构的思想和工具的使用，更重要的是，本书包含了商业智能的规划思想和原理，也渗透了架构过程中的难点和解决方案。本书并不拘泥于细节部分，但是试图帮助读者深入理解商业智能的核心技术，包括相关的理论、规划思想和实际案例等。

希望本书能够带您进入到真正的数据世界中。在数据的海洋中自由地发挥，不断地创造价值。

本书涵盖的内容

本书涵盖的内容可以分成如下几个部分。

理论篇：主要包括商业智能概述、数据仓库理论知识、数据挖掘和分析、大数据理论知识等内容。

项目篇：主要包括商业智能项目需求的定义、商业智能项目模型的建立、商业智能数据仓库系统应用实例、大数据架构与实践、电力行业和金融行业的商业智能案例等内容。

工具篇：详细介绍两大商业智能工具 Cognos 与 Informatica 的理论知识和使用方法。

实践篇：包括 Cognos 报表的开发、部署和实践。

本书附赠部分源代码和一些有价值的文档模板。

阅读背景

目前，很多企业引入了商业智能。商业智能已经成为了企业信息化的重要的解决方案和技术手段。从零售企业到金融机构，从电力行业到环保产业，商业智能正在引起一场商业分析和战略决策的革命。预计，很快就会有超过半数的企业使用商业智能技术去开拓和抢占市场了。

小时候曾经有一本书给我留下了深刻的印象，大概内容是教小学生如何写作文的，包括如何写议论文、记叙文、散文等。书中并没有很刻板地去阐述写作的步骤，而是通过主人公（小胖）的一个个小故事，把一些看似偶然，却又非常合理的情节，包括在公园、火车站、旅游景点等场景下，以对话的形式将写作的方法一一告诉读者，语言风趣幽默。身为读者的我读后有了相当大的收获，尽管已时隔几十年，仍然让我不能忘怀。本书的内容，虽然不能以故事情节或者对话的形式展现给大家，但是也力求不用晦涩难懂的语言去解释概念，而是用简单易懂的例子去说明问题，以求既能让初学者快速进入到商业智能的世界中，也可让专业人士在轻松阅读的过程中有所收益。

作者

内容简介

目录

致 谢

我要感谢帮助我完成这本书的所有人。

首先感谢刘国峰对本书的重要贡献，他为本书的修订付出了心血和努力。

感谢公司的同事们，他们都以各种方式为本书的修订做出了贡献。

最后，也是最重要的，我要感谢亲人给予我的不懈支持，我的母亲（张丽华）、父亲（王贵林），还有我的岳父（丁一贤）、岳母（赵桂荣），特别是我的妻子（丁玲玲）和2岁的女儿（王预萱），他们给予了我最大的精神支持和帮助。如果没有他们，我很难想象能完成这本书。

王 飞

目 录

前言
致谢

理 论 篇

第 1 章 商业智能简介	1
1.1 商业智能概述	1
1.1.1 商业智能的定义	1
1.1.2 商业智能的作用	3
1.1.3 商业智能的处理过程	3
1.1.4 商业智能的功能	3
1.1.5 商业智能的发展趋势	4
1.2 关于商业智能的核心技术	5
1.2.1 数据仓库	5
1.2.2 数据挖掘和分析	5
1.2.3 ETL 处理技术	5
1.2.4 联机分析处理 (OLAP) 技术	5
1.2.5 可视化分析	5
1.2.6 大数据技术	6
1.2.7 商业智能元数据管理	6
1.3 商业智能的实施方法和步骤	6
1.3.1 商业智能的实施方法	7
1.3.2 商业智能的实施步骤	7
1.4 商业智能项目成功的关键	8
1.5 商业智能项目的组织机构	8
1.6 本章小结	9
第 2 章 数据仓库的理论知识	11
2.1 数据仓库概述	11
2.1.1 数据仓库产生的背景	11
2.1.2 数据仓库定义	11
2.1.3 数据仓库和商业智能之间的关系	12
2.2 数据仓库的特点	12
2.2.1 面向主题	12
2.2.2 集成性	13
2.2.3 稳定性	13
2.2.4 反映历史变化	13

2.3	数据仓库的优势	14
2.4	数据仓库和数据库的区别	14
2.5	数据仓库开发过程介绍	14
2.5.1	规划分析阶段	14
2.5.2	设计实现阶段	15
2.5.3	使用维护阶段	15
2.6	数据仓库系统组成部分介绍	17
2.6.1	数据源分析	17
2.6.2	数据迁移	18
2.6.3	选择数据的存储结构	19
2.6.4	元数据	19
2.7	数据仓库模型设计介绍	20
2.7.1	概念模型	20
2.7.2	逻辑模型	21
2.7.3	物理模型	23
2.8	数据集市介绍	24
2.8.1	数据集市概述	24
2.8.2	数据集市和数据仓库的联系和区别	26
2.8.3	数据集市的目标分析	27
2.8.4	数据集市的技术特性	27
2.9	ODS 介绍	28
2.9.1	ODS 的概述	28
2.9.2	ODS 系统与数据库系统、数据仓库系统的区别	32
2.9.3	基于 ODS 和基于数据仓库的 OLAP 之间的关系	33
2.9.4	ODS 系统的功能	33
2.9.5	ODS 系统的架构	34
2.10	数据仓库实施详细步骤	34
2.10.1	需求分析	34
2.10.2	数据仓库的逻辑分析	35
2.10.3	设计 ODS 系统	35
2.10.4	数据仓库建模	37
2.10.5	数据集市建模	38
2.10.6	数据源分析	38
2.10.7	数据的获取与整合	38
2.10.8	应用分析	38
2.10.9	报表展现	39
2.11	在大数据环境下的数据仓库的建设	39
2.12	数据仓库建设路线图	40
2.13	数据仓库的作用	41

2.14	数据仓库的建设意义	42
2.15	本章小结	42
第3章	数据挖掘和分析理论知识	44
3.1	什么是数据挖掘	44
3.2	数据挖掘方法的几个步骤	44
3.3	数据挖掘常用算法	45
3.4	数据仓库和数据挖掘之间的关系	45
3.5	数据挖掘的主要过程	45
3.6	数据挖掘的主要应用——客户精准营销	46
3.6.1	客户精准营销背景	46
3.6.2	关于旅游行业的客户精准营销	46
3.6.3	关于银行业的客户精准营销	49
3.7	本章小结	50
第4章	商业智能 ETL 理论知识	51
4.1	ETL 在数据仓库中的重要地位	51
4.2	ETL 的一般过程	52
4.3	研究 ETL 的本质	52
4.4	ETL 的体系结构	54
4.5	ETL 的难点	54
4.6	主流的 ETL 工具	54
4.7	ETL 的作用	55
4.8	详解 ETL 过程	55
4.8.1	数据抽取	56
4.8.2	数据清洗	56
4.8.3	数据转换	57
4.8.4	数据加载	58
4.8.5	ETL 的日志	59
4.9	ETL 优化	59
4.10	ETL 设计规范要点	59
4.11	ETL 的框架结构	60
4.12	ETL 的实施策略	61
4.13	本章小结	62
第5章	商业智能联机分析处理理论简介	64
5.1	OLAP 介绍	64
5.2	OLAP 系统与 OLTP 系统的区别	65
5.3	OLAP 的实现方法	65
5.4	OLAP 的基本目标和特点	67
5.5	建立 OLAP 的过程	67
5.6	OLAP 与数据仓库的关系	68

5.7	OLAP 系统的实施过程	68
5.8	OLAP 模型的设计与实现	69
5.9	本章小结	70
第 6 章	数据可视化分析理论知识	71
6.1	什么是数据可视化分析	71
6.2	数据可视化的表现形式	72
6.3	本章小结	73
第 7 章	大数据理论知识	74
7.1	大数据概念的提出	74
7.2	什么是大数据?	75
7.3	大数据处理技术	75
7.4	大数据应用案例	76
7.5	大数据和传统商业智能的结合	76
7.6	本章小结	76
第 8 章	商业智能元数据理论知识	77
8.1	元数据的定义	77
8.2	元数据的重要性	78
8.3	元数据的类型	78
8.4	元数据的作用	78
8.5	元数据的管理	79
8.6	元数据包含的内容	80
8.7	本章小结	80

目 录

第 9 章	商业智能项目需求的定义	81
9.1	商业智能项目的启动	81
9.2	商业智能项目的需求定义	84
9.3	系统原型的建立	90
9.4	验收和评审的内容	90
9.5	本章小结	91
第 10 章	商业智能项目模型的建立	92
10.1	数据模型的设计原则	92
10.2	企业模型的意义	92
10.2.1	企业模型的定义	92
10.2.2	建设企业模型的意义	92
10.2.3	企业数据模型和其他模型之间的关系	93
10.2.4	与企业数据模型相关的概念	93
10.2.5	企业数据模型转换到数据仓库模型的步骤	94
10.3	概念模型的设计	96

10.4	逻辑模型的设计	97
10.4.1	ODS 逻辑模型	97
10.4.2	数据仓库逻辑模型	101
10.4.3	数据集市逻辑模型	106
10.5	物理模型的设计	109
10.5.1	ODS 物理模型的设计	109
10.5.2	数据仓库物理模型的设计	110
10.5.3	数据集市物理模型的设计	112
10.6	本章小结	116
第 11 章	商业智能数据仓库系统应用实例	118
11.1	定义数据仓库项目的生命周期	118
11.2	数据仓库粒度的划分	119
11.3	企业辅助决策分析系统的构建	119
11.4	决策分析系统一般的部署方案和步骤	120
11.4.1	提供系统安装软件的体系结构	120
11.4.2	部署系统的数据库环境	121
11.4.3	ETL 环境的部署	122
11.4.4	报表展示环境的部署	122
11.5	数据仓库建设的难点	122
11.6	本章小结	123
第 12 章	商业银行数据仓库建设规划	124
12.1	商业银行数据仓库建设概况	124
12.2	目前国内商业银行数据仓库面临的瓶颈	124
12.3	商业银行数据仓库建设面临哪些问题	125
12.4	商业银行数据仓库建设思路及系统情况	126
12.4.1	某商业银行建设数据仓库时遇到的挑战	126
12.4.2	某商业数据仓库架构存在的问题	127
12.4.3	对该行数据仓库目标架构的建议	128
12.5	商业银行数据仓库建设启示	129
12.6	本章小结	130
第 13 章	电力行业数据仓库的建设规划	131
13.1	电力行业数据仓库建设难点	131
13.2	电力行业数据仓库逻辑架构	132
13.3	电力行业数据仓库能力蓝图	132
13.4	数据仓库促进电力业务的发展	133
13.5	数据仓库建设策略比较	134
13.6	电力行业数据仓库模型建立过程	134
13.7	电力行业数据仓库的架构设计	134
13.8	本章小结	136

第 14 章 商业智能项目规划和管理	138
14.1 项目团队的组织结构	138
14.2 项目角色划分及技能要求	139
14.3 定义领导组的职责和主要任务	140
14.4 如何定义商业智能项目的进度	140
14.5 如何定义商业智能项目的过程	141
14.6 本章小结	142
第 15 章 商业智能应用介绍	143
15.1 商业智能应用设计的原则	143
15.2 商业智能应用的实施步骤	144
15.3 商业智能具有的应用功能	144
15.4 商业智能应用实例	145
15.5 本章小结	146
第 16 章 数据抽取、转换和加载	147
16.1 ETL 的定义和总体架构	147
16.2 ETL 的加载方法	150
16.2.1 以时间戳作为加载条件	150
16.2.2 利用源表的日志信息对目标表进行数据加载	153
16.2.3 通过全表对比的方式进行数据加载	153
16.2.4 全表删除后再进行数据加载的方式	156
16.3 利用 ETL 构建企业级数据仓库	157
16.4 ETL 的设计过程	158
16.5 ETL 的备份与恢复	160
16.5.1 数据的备份	160
16.5.2 数据备份存放的介质以及目录结构	160
16.5.3 ETL 程序的备份	161
16.5.4 数据的恢复方案	161
16.6 ETL 数据质量的改进	163
16.6.1 ETL 数据质量分析	163
16.6.2 ETL 数据质量改进的方法和目标	163
16.6.3 推动 ETL 数据质量改进的方法	163
16.6.4 ETL 的技术路线选择	164
16.7 ETL 应用举例	165
16.7.1 ETL 分析需求	165
16.7.2 ETL 数据源说明	166
16.7.3 ODS 设计与抽取	166
16.7.4 数据仓库 (DW) 设计与抽取	169
16.7.5 数据集市 (DM) 设计与抽取	171
16.8 本章小结	173

第 17 章 联机分析处理	175
17.1 OLAP 的概念	175
17.2 OLAP 的实施	175
17.2.1 建立“维”的概念	175
17.2.2 多维分析技术	176
17.2.3 OLAP 实施的一般过程	178
17.3 OLAP 的应用实例	179
17.3.1 案例背景	179
17.3.2 需求	180
17.3.3 数据准备	180
17.3.4 浏览分析数据	181
17.4 OLAP 系统设计的一般步骤	182
17.5 本章小结	183
第 18 章 应用举例	185
18.1 项目工作计划的制订	185
18.1.1 对项目背景与目的的描述	185
18.1.2 确定项目需要交付的成果	185
18.1.3 制定项目管理文档	186
18.1.4 项目进度划分	187
18.2 需求分析	188
18.2.1 业务需求	188
18.2.2 功能需求	189
18.3 营销系统设计	191
18.3.1 总体架构设计	191
18.3.2 营销辅助决策系统的 ETL 架构设计	192
18.3.3 营销辅助决策系统数据访问功能设计	193
18.3.4 营销辅助决策系统展示方式设计	193
18.3.5 营销辅助决策系统主题分析功能设计	194
18.3.6 数据模型设计	195
18.4 ETL 数据抽取	203
18.4.1 ETL 物理设计	203
18.4.2 从源数据库抽取到 ODS 数据缓冲区	204
18.4.3 从 ODS 数据缓冲区抽取到 ODS 统一信息视图区	204
18.4.4 从 ODS 统一信息视图区抽取到数据仓库	205
18.4.5 从数据仓库抽取到数据集市	205
18.5 报表展示	206
18.6 编写测试报告	206
18.7 编写用户手册	208
18.8 软件发布	209

18.9	系统运行维护	209
18.10	本章小结	210
第 19 章	大数据架构与实践	211
19.1	大数据概述	211
19.2	大数据的处理技术之一——流数据	212
19.3	大数据下的数据架构	212
19.4	大数据在银行业的应用与实践	213
19.5	本章小结	216
第 20 章	金融行业的商业智能概述	217
20.1	金融行业实施商业智能的背景	217
20.2	商业智能在金融行业的作用	217
20.3	金融行业实施商业智能的措施	218
20.4	本章小结	218
第 21 章	电力行业商业智能概述	219
21.1	电力行业面临的挑战	219
21.2	建设企业级数据仓库的原因	219
21.3	电力行业数据仓库的执行架构	220
21.4	数据仓库开发的阶段、任务和流程	220
21.5	数据仓库运维内容	221
21.6	电力行业数据仓库的建设方法	221
21.7	本章小结	224

工 具 篇

第 22 章	Informatica PowerCenter 工具简介	225
22.1	Informatica PowerCenter 介绍	225
22.1.1	Informatica 的特点	226
22.1.2	Informatica 的优势	226
22.2	Informatica PowerCenter 工具概况	227
22.3	Informatica Servers 引擎	227
22.4	Administration Console	228
22.4.1	登录方式	229
22.4.2	相关术语	230
22.5	PowerCenter Designer	230
22.5.1	菜单	231
22.5.2	工具栏	233
22.5.3	导航	235
22.5.4	工作区	236
22.5.5	输出窗口	237
22.6	Repository Manager	237

22.6.1	菜单	238
22.6.2	工具栏	239
22.6.3	导航	239
22.6.4	工作区	239
22.7	Workflow Manager	240
22.7.1	菜单	240
22.7.2	工具栏	241
22.7.3	导航	242
22.7.4	工作区	243
22.7.5	输出窗口	243
22.8	Workflow Monitor	243
22.8.1	工具栏	244
22.8.2	监控区	245
22.8.3	属性	246
22.9	本章小结	246
第 23 章	Cognos 工具简介	247
23.1	Cognos 介绍	247
23.1.1	Cognos 的历史	247
23.1.2	Cognos 的特点	247
23.1.3	Cognos 的现状	248
23.2	Cognos 工具浏览	248
23.3	Cognos Configuration	249
23.4	Cognos Connection	250
23.4.1	菜单功能	250
23.4.2	选项卡介绍	251
23.4.3	工具栏的使用	251
23.5	Query Studio	251
23.6	Analysis Studio 介绍	256
23.6.1	Analysis Studio 的组成	257
23.6.2	菜单功能	257
23.7	Report Studio 介绍	258
23.7.1	Report Studio 的组成	259
23.7.2	菜单功能	262
23.8	Event Studio 介绍	264
23.8.1	Event 介绍	264
23.8.2	Task 介绍	265
23.9	Framework Manager 建模工具	267
23.9.1	菜单介绍	268
23.9.2	工作区	271

23.10 Transformer 介绍	274
23.10.1 Data Sources	274
23.10.2 Dimension Map	275
23.10.3 Measures	276
23.10.4 PowerCubes	276
23.10.5 Customer Views	276
23.10.6 Signon	277
23.11 Cognos PowerPlay	277
23.11.1 菜单介绍	278
23.11.2 工具栏	280
23.11.3 维度视图	280
23.12 本章小结	280

实 践 篇

第 24 章 Informatica 的安装与快速入门	282
24.1 Informatica PowerCenter 的安装方案	282
24.1.1 安装前的准备	282
24.1.2 服务器端安装	283
24.1.3 客户端安装	291
24.2 Informatica PowerCenter 工具的一般使用流程	295
24.3 Informatica PowerCenter 快速入门	296
24.3.1 前期准备	296
24.3.2 定义源数据	301
24.3.3 定义目标数据	306
24.3.4 定义映射规则	306
24.3.5 定义 workflow	309
24.3.6 启动 workflow	316
24.4 本章小结	317
第 25 章 Informatica 实例	319
25.1 缓慢变化维的概念	319
25.2 缓慢变化维的实现方式	319
25.2.1 覆盖方式	319
25.2.2 全历史记录	324
25.2.3 标记位方式	328
25.2.4 时间戳方式	333
25.2.5 记录最新记录和上一次历史	337
25.3 本章小结	341
第 26 章 Cognos 安装与快速入门	342
26.1 Cognos 8 安装	342

26.1.1	Cognos 服务器安装	342
26.1.2	Cognos Framework 安装	345
26.1.3	Cognos Transformer 安装	347
26.1.4	Cognos 语言包安装	349
26.1.5	Cognos PowerPlay 安装	354
26.2	Cognos 入门前的准备工作	358
26.2.1	创建 Oracle 数据库实例	358
26.2.2	执行建表空间和创建数据库用户的脚本	363
26.2.3	Cognos 部署设置	364
26.2.4	配置 Cognos 服务	364
26.2.5	启动 Cognos 服务	369
26.3	Cognos 入门实例一	369
26.3.1	使用 Framework 建模	371
26.3.2	使用 Report Studio 制作报表	382
26.3.3	查看报表	389
26.4	Cognos 入门实例二	390
26.4.1	使用 Framework 建模	390
26.4.2	使用 Transformer 立方体多维建模	393
26.4.3	使用 Analysis Studio 制作多维报表	405
26.4.4	查看报表	408
26.5	本章小结	409
第 27 章	Cognos 实例	410
27.1	Section 报表的开发	410
27.2	Conditional Style 报表的开发	413
27.3	图表型报表的开发	419
27.4	Drill-Through 报表的开发	425
27.5	本章小结	434
第 28 章	Cognos 的安全管理	435
28.1	Cognos 安全性介绍	435
28.1.1	Cognos 应用防火墙	435
28.1.2	身份验证	436
28.1.3	访问授权	437
28.1.4	加密服务	438
28.2	Cognos 安全部署	439
28.2.1	操作系统安全技术	439
28.2.2	网络安全技术	441
28.2.3	应用安全技术	441
28.2.4	安全审计	442
28.3	Cognos 安全实践	443