

内 容 简 介

本书是由全国计算机专业技术资格考试办公室组织编写的考试用书。本书根据《系统集成项目管理工程师考试大纲》(2023年审定通过)编写,阐述了系统集成项目管理工程师(项目经理)岗位所要求的主要知识以及应用技术。

本书主要内容包括信息化发展、信息技术发展、信息技术服务、信息系统架构、软件工程、数据工程、软硬件系统集成、信息安全工程、项目管理概论、启动过程组、规划过程组、执行过程组、监控过程组、收尾过程组、组织保障、监理基础知识、法律法规和标准规范、职业道德规范。

本书是参加系统集成项目管理工程师考试应试者的必读教材,也可以作为信息化教育的培训或辅导用书,还可以作为高等院校信息管理专业的教学或参考用书。由于书中提供的一些技术、工具和方法具有较强的实践性,本书也能够作为在职人员的工具书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。举报:010-62782989, beiqinquan@tup.tsinghua.edu.cn。

图书在版编目(CIP)数据

系统集成项目管理工程师教程 / 张树玲, 宋跃武主编. —3版. —北京: 清华大学出版社, 2024.1

全国计算机技术与软件专业技术资格(水平)考试指定用书

ISBN 978-7-302-65219-9

I. ①系… II. ①张… ②宋… III. ①系统集成技术—项目管理—资格考试—自学参考资料
IV. ① TP311.5

中国国家版本馆 CIP 数据核字 (2024) 第 003736 号

责任编辑: 杨如林

封面设计: 杨玉兰

责任校对: 胡伟民

责任印制: 宋 林

出版发行: 清华大学出版社

网 址: <https://www.tup.com.cn>, <https://www.wqxuetang.com>

地 址: 北京清华大学学研大厦 A 座

邮 编: 100084

社 总 机: 010-83470000

邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者: 河北鹏润印刷有限公司

经 销: 全国新华书店

开 本: 185mm×230mm

印 张: 38.25 防伪页: 1

字 数: 1200 千字

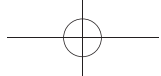
版 次: 2009 年 3 月第 1 版

2024 年 1 月第 3 版

印 次: 2024 年 1 月第 1 次印刷

定 价: 139.00 元

产品编号: 103164-01



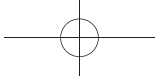
前言

计算机技术与软件专业技术资格（水平）考试（以下简称“软考”）中的系统集成项目管理工程师资格自 2005 年开考以来，已经培养了大批项目管理人员，这些专业技术人才在信息系统建设、服务管理与应用、运维和创新等方面发挥了重要作用，为保证组织的信息系统集成建设质量，提高信息系统项目开发绩效，推进组织转型升级等做出了贡献。

采用现代管理理论作为计划、设计、控制的方法论，将计算机系统、数据库、网络、安全设施、应用软件等资源按照规划的结构和秩序，有机地整合到一个边界清晰、敏捷高效的信息系统中，以达到预定的系统目标，这是系统集成的主要工作。综合运用相关知识、技能、工具和技术在一定的时间、成本、质量等要求下，为实现预定的系统目标而进行的管理计划、设计、开发、实施、运维等方面的活动称为信息系统项目管理。实施项目管理的项目管理工程师（项目经理）和项目管理师（高级项目经理）岗位已经成为组织信息化建设和数字化转型过程中不可缺少的重要岗位，对这个岗位的人才选拔和评价是信息行业人才队伍建设的重要组成部分。

在信息技术的推动下，人类社会正在加速进入全新发展时期，以数据要素开发利用为基础的数字时代、数字经济快速发展，以人工智能、区块链和物联网等为代表的新一代信息技术，正在深刻影响着各行业领域的改革创新与转型升级，信息系统相关技术、建设、管理、运行等数字化能力需求持续强化，系统集成项目管理将面临新的行业背景和形势。2023 年，为适应经济与社会的发展需要以及信息技术和项目管理技术的发展现状，全国计算机专业技术资格考试办公室广泛吸纳当前最新的研究成果，在原大纲的基础上，组织专家对系统集成项目管理工程师考试大纲进行了较大幅度的修改，增强了信息化发展、信息技术、信息技术服务、系统集成工程建设等方面的知识要求，以实用性和适用性为主线，并结合敏捷和价值驱动等，丰富和完善了项目的知识体系要求，更新、丰富和完善了系统集成项目相关标准与法律法规的要求。

依据新修订的《系统集成项目管理工程师考试大纲》（2023 年审定通过），全国计算机专业技术资格考试办公室组织专家对《系统集成项目管理工程师教程》（以下简称《教程》）的第 2 版进行了修订。修订的第 3 版《教程》在论述项目管理知识体系时尽量做到体例一致，叙述方式及逻辑一致，专业用语一致；论述的方法、工具和技术，与实际工作开展结合紧密，可以直接在系统集成项目管理的实践中运用。考虑系统集成建设与各行业领域发展深度融合的需求，系统集成项目管理工程师具有体系化的信息系统发展知识、系统化的信息技术知识、全面的信息技术服务知识、立体的系统集成工程知识、深入的信息系统项目新型管理知识等，对项目的成功无疑是至关重要的。因此，第 3 版《教程》基于国家相关发展战略，结合技术发展新趋势及应用新场景，丰富了信息技术的基础知识及其开发应用趋势；基于组织系统集成发展与项目的高度融合性，丰富了信息技术服务和各类系统工程相关知识；基于新时期对系统集成项目管理的敏捷性与价值需求，重新梳理了项目管理相关知识。考虑各类信息系统项目实践的差异化和个性化，信息系统建设与组织发展的深度融合强化，以及各类网络化知识的丰富，修订教程



II 系统集成项目管理工程师教程（第3版）

适度给出了知识示例相关材料。

另外，为了便于参加软考的专业技术人员复习应考，第3版《教程》在每章还提供了相应练习题。

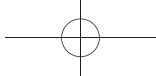
第1章信息化发展由宋跃武、刘媛媛编写；第2章信息技术发展由宋俊典、王铮、宋跃武编写；第3章信息技术服务由李美翠、曾寰嵩、王林编写；第4章信息系统架构由孙佩、李美翠、张荣静编写；第5章软件工程由李超、白文江、王林编写；第6章数据工程由宋俊典、刘媛媛、郝孝宇编写；第7章软硬件系统集成由刘瑞慧、隋志巍、李超、崔晓柳编写；第8章信息安全工程由林程远、隋志巍、彭高编写；第9章项目管理概论由刘玲、宋跃武编写；第10章启动过程组由李少杰、刘玲、宋跃武编写；第11章规划过程组由李京、刘玲、宋跃武编写；第12章执行过程组由唐百惠、张树玲编写；第13章监控过程组由刘娜、张树玲、刘玲编写；第14章收尾过程组由李少杰、张树玲、李静编写；第15章组织保障由尹正茹、岳素林、郭爽编写；第16章监理基础知识由宋跃武、贾卓生、李京编写；第17章法律法规和标准规范由李修仪、朱盼盼、岳素林编写；第18章职业道德规范由岳素林、刘玲、李世喆编写。

另外，张树玲和宋跃武依据考试大纲对全书做了内容统筹、章节结构设计和统稿，岳素林、李超等参加了本书的策划以及部分章节的审校，刘玲和李超提供了各类实践经验并对相关内容进行了提炼总结，李美翠参与了本书部分章节的审校工作。清华大学出版社为本书的编写做了大量的组织管理工作，在此表示感谢。

由于编者水平所限，书中难免有不当之处，恳请读者不吝赐教并提出宝贵意见，相信读者的反馈将会为未来的再次修订提供良好的帮助。

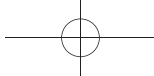
编 者

2023 年 12 月



目 录

第 1 章 信息化发展	1	2.2.1 物联网.....	76
1.1 信息与信息化.....	1	2.2.2 云计算.....	77
1.1.1 信息基础.....	2	2.2.3 大数据.....	81
1.1.2 信息系统基础.....	4	2.2.4 区块链.....	83
1.1.3 信息化基础.....	7	2.2.5 人工智能.....	87
1.2 现代化基础设施.....	12	2.2.6 虚拟现实.....	90
1.2.1 新型基础设施建设.....	12	2.3 新一代信息技术发展展望.....	91
1.2.2 工业互联网.....	13	2.4 本章练习.....	93
1.2.3 城市物联网.....	16	第 3 章 信息技术服务	95
1.3 产业现代化.....	19	3.1 内涵与外延.....	95
1.3.1 农业农村现代化.....	19	3.1.1 服务的特征.....	95
1.3.2 工业现代化.....	20	3.1.2 IT服务的内涵.....	96
1.3.3 服务现代化.....	23	3.1.3 IT服务的外延.....	96
1.4 数字中国.....	25	3.1.4 IT服务业的特征.....	97
1.4.1 数字经济.....	25	3.2 原理与组成.....	97
1.4.2 数字政府.....	30	3.2.1 IT服务原理.....	98
1.4.3 数字社会.....	34	3.2.2 组成要素.....	98
1.4.4 数字生态.....	39	3.3 服务生命周期.....	102
1.5 数字化转型与元宇宙.....	41	3.3.1 战略规划.....	103
1.5.1 数字化转型.....	41	3.3.2 设计实现.....	105
1.5.2 元宇宙.....	47	3.3.3 运营提升.....	107
1.6 本章练习.....	49	3.3.4 退役终止.....	109
第 2 章 信息技术发展	50	3.4 服务标准化.....	110
2.1 信息技术及其发展.....	50	3.4.1 服务产业化.....	110
2.1.1 计算机软硬件.....	50	3.4.2 服务标准.....	112
2.1.2 计算机网络.....	52	3.5 服务质量评价.....	115
2.1.3 存储和数据库.....	59	3.5.1 相关方模型.....	115
2.1.4 信息安全.....	66	3.5.2 互动模型.....	116
2.1.5 信息技术的发展.....	75	3.5.3 质量模型.....	117
2.2 新一代信息技术及应用.....	76	3.6 服务发展.....	118
		3.6.1 发展环境.....	118



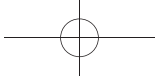
IV 系统集成项目管理工程师教程（第3版）

3.6.2 发展现状与趋势.....	119
3.7 服务集成与实践.....	121
3.7.1 实践背景.....	122
3.7.2 服务产品与组合.....	125
3.7.3 项目组织与里程碑计划.....	129
3.7.4 项目风险识别与控制.....	132
3.8 本章练习.....	134
第4章 信息系统架构	136
4.1 架构基础.....	136
4.1.1 指导思想.....	136
4.1.2 设计原则.....	136
4.1.3 建设目标.....	137
4.1.4 总体框架.....	138
4.2 系统架构.....	140
4.2.1 架构定义.....	140
4.2.2 架构分类.....	141
4.2.3 一般原理.....	143
4.2.4 常用架构模型.....	144
4.2.5 规划与设计.....	147
4.2.6 价值驱动的体系结构.....	153
4.3 应用架构.....	156
4.3.1 基本原则.....	156
4.3.2 分层分组.....	156
4.4 数据架构.....	158
4.4.1 发展演进.....	158
4.4.2 基本原则.....	159
4.4.3 架构举例.....	160
4.5 技术架构.....	161
4.5.1 基本原则.....	161
4.5.2 架构举例.....	162
4.6 网络架构.....	163
4.6.1 基本原则.....	163
4.6.2 局域网架构.....	163
4.6.3 广域网架构.....	166
4.6.4 移动通信网架构.....	170
4.6.5 软件定义网络.....	172

4.7 安全架构.....	172
4.7.1 安全威胁.....	173
4.7.2 定义和范围.....	174
4.7.3 整体架构设计.....	175
4.7.4 网络安全架构设计.....	179
4.7.5 数据库系统安全设计.....	184
4.7.6 安全架构设计案例分析.....	187
4.8 云原生架构.....	189
4.8.1 发展概述.....	189
4.8.2 架构定义.....	191
4.8.3 基本原则.....	192
4.8.4 常用架构模式.....	194
4.8.5 云原生案例.....	197
4.9 本章练习.....	200

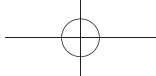
第5章 软件工程..... 202

5.1 软件工程定义.....	202
5.2 软件需求.....	202
5.2.1 需求的层次.....	203
5.2.2 质量功能部署.....	203
5.2.3 需求获取.....	204
5.2.4 需求分析.....	204
5.2.5 需求规格说明书.....	208
5.2.6 需求变更.....	210
5.2.7 需求跟踪.....	212
5.3 软件设计.....	212
5.3.1 结构化设计.....	213
5.3.2 面向对象设计.....	216
5.3.3 统一建模语言.....	217
5.3.4 设计模式.....	220
5.4 软件实现.....	220
5.4.1 软件配置管理.....	220
5.4.2 软件编码.....	221
5.4.3 软件测试.....	222
5.5 部署交付.....	225
5.5.1 软件部署.....	225
5.5.2 软件交付.....	226



5.5.3	持续交付.....	226
5.5.4	持续部署.....	227
5.5.5	部署和交付的新趋势.....	228
5.6	软件质量管理.....	228
5.7	软件过程能力成熟度.....	230
5.7.1	成熟度模型.....	230
5.7.2	成熟度等级.....	231
5.8	本章练习.....	232
第 6 章 数据工程		234
6.1	数据采集和预处理.....	234
6.1.1	数据采集.....	234
6.1.2	数据预处理.....	235
6.1.3	数据预处理方法.....	235
6.2	数据存储及管理.....	236
6.2.1	数据存储.....	236
6.2.2	数据归档.....	238
6.2.3	数据备份.....	238
6.2.4	数据容灾.....	240
6.3	数据治理和建模.....	241
6.3.1	元数据.....	241
6.3.2	数据标准化.....	242
6.3.3	数据质量.....	243
6.3.4	数据模型.....	244
6.3.5	数据建模.....	245
6.4	数据仓库和数据资产.....	246
6.4.1	数据仓库.....	246
6.4.2	主题库.....	247
6.4.3	数据资产管理.....	247
6.4.4	数据资源编目.....	248
6.5	数据分析及应用.....	249
6.5.1	数据集成.....	249
6.5.2	数据挖掘.....	251
6.5.3	数据服务.....	253
6.5.4	数据可视化.....	254
6.6	数据脱敏和分类分级.....	255
6.6.1	数据脱敏.....	255

6.6.2	数据分类.....	256
6.6.3	数据分级.....	257
6.7	本章练习.....	257
第 7 章 软硬件系统集成		259
7.1	系统集成基础.....	259
7.2	基础设施集成.....	261
7.2.1	弱电工程.....	261
7.2.2	网络集成.....	263
7.2.3	数据中心集成.....	264
7.3	软件集成.....	266
7.3.1	基础软件集成.....	266
7.3.2	应用软件集成.....	271
7.3.3	其他软件集成.....	275
7.4	业务应用集成.....	276
7.5	本章练习.....	278
第 8 章 信息安全工程		279
8.1	信息安全管理.....	279
8.1.1	保障要求.....	279
8.1.2	管理内容.....	280
8.1.3	管理体系.....	285
8.1.4	等级保护.....	287
8.2	信息安全系统.....	290
8.2.1	安全机制.....	291
8.2.2	安全服务.....	292
8.3	工程体系架构.....	293
8.3.1	安全工程基础.....	293
8.3.2	ISSE-CMM基础.....	295
8.3.3	ISSE过程.....	295
8.3.4	ISSE-CMM体系结构.....	298
8.4	本章练习.....	301
第 9 章 项目管理概论		303
9.1	PMBOK 的发展.....	303
9.2	项目基本要素.....	304
9.2.1	项目基础.....	304



VI 系统集成项目管理工程师教程（第3版）

9.2.2	项目管理	306
9.2.3	项目成功的标准	307
9.2.4	项目、项目集、项目组合和运营管理之间的关系	308
9.2.5	项目运行环境	312
9.2.6	组织系统	314
9.2.7	项目管理和产品管理	317
9.3	项目经理的角色	318
9.4	项目生命周期和项目阶段	318
9.4.1	定义与特征	318
9.4.2	生命周期类型	320
9.5	项目立项管理	322
9.5.1	项目建议与立项申请	322
9.5.2	项目可行性研究	323
9.5.3	项目评估与决策	331
9.6	项目管理过程组	333
9.7	项目管理原则	336
9.8	项目管理知识领域	349
9.9	价值交付系统	351
9.10	本章练习	353

第10章 启动过程组 354

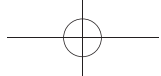
10.1	制定项目章程	355
10.1.1	主要输入	356
10.1.2	主要输出	357
10.2	识别干系人	358
10.2.1	主要输入	360
10.2.2	主要工具与技术	361
10.2.3	主要输出	362
10.3	启动过程组的重点工作	362
10.3.1	项目启动会议	362
10.3.2	关注价值和目标	363
10.4	本章练习	364

第11章 规划过程组 365

11.1	制订项目管理计划	367
11.1.1	主要输入	368

11.1.2	主要输出	369
11.2	规划范围管理	372
11.2.1	主要输入	372
11.2.2	主要输出	373
11.3	收集需求	373
11.3.1	主要输入	374
11.3.2	主要工具与技术	375
11.3.3	主要输出	377
11.4	定义范围	378
11.4.1	主要输入	379
11.4.2	主要输出	379
11.5	创建 WBS	380
11.5.1	主要输入	380
11.5.2	主要工具与技术	381
11.5.3	主要输出	383
11.6	规划进度管理	384
11.6.1	主要输入	384
11.6.2	主要输出	385
11.7	定义活动	385
11.7.1	主要输入	386
11.7.2	主要工具与技术	386
11.7.3	主要输出	386
11.8	排列活动顺序	387
11.8.1	主要输入	387
11.8.2	主要工具与技术	388
11.8.3	主要输出	391
11.9	估算活动持续时间	391
11.9.1	主要输入	393
11.9.2	主要工具与技术	393
11.9.3	主要输出	394
11.10	制订进度计划	395
11.10.1	主要输入	396
11.10.2	主要工具与技术	397
11.10.3	主要输出	402
11.11	规划成本管理	404
11.11.1	主要输入	404
11.11.2	主要输出	405

11.12 估算成本.....	405	11.22 规划风险应对.....	445
11.12.1 主要输入.....	406	11.22.1 主要输入.....	446
11.12.2 主要输出.....	407	11.22.2 主要工具与技术.....	447
11.13 制定预算.....	408	11.22.3 主要输出.....	449
11.13.1 主要输入.....	408	11.23 规划采购管理.....	449
11.13.2 主要输出.....	409	11.23.1 主要输入.....	451
11.14 规划质量管理.....	410	11.23.2 主要输出.....	451
11.14.1 主要输入.....	411	11.23.3 合同类型.....	454
11.14.2 主要工具与技术.....	411	11.23.4 合同内容.....	458
11.14.3 主要输出.....	414	11.24 规划干系人参与.....	459
11.15 规划资源管理.....	417	11.24.1 主要输入.....	460
11.15.1 主要输入.....	418	11.24.2 主要工具与技术.....	460
11.15.2 主要工具与技术.....	418	11.24.3 主要输出.....	461
11.15.3 主要输出.....	419	11.25 本章练习.....	462
11.16 估算活动资源.....	420	第 12 章 执行过程组.....	465
11.16.1 主要输入.....	421	12.1 指导与管理项目工作.....	466
11.16.2 主要输出.....	422	12.1.1 主要输入.....	467
11.17 规划沟通管理.....	422	12.1.2 主要输出.....	468
11.17.1 主要输入.....	423	12.2 管理项目知识.....	470
11.17.2 主要工具与技术.....	424	12.2.1 主要输入.....	472
11.17.3 主要输出.....	425	12.2.2 主要输出.....	472
11.18 规划风险管理.....	425	12.3 管理质量.....	472
11.18.1 风险基本概念.....	426	12.3.1 主要输入.....	474
11.18.2 主要输入.....	429	12.3.2 主要工具与技术.....	474
11.18.3 主要输出.....	429	12.3.3 主要输出.....	477
11.19 识别风险.....	432	12.4 获取资源.....	478
11.19.1 主要输入.....	433	12.4.1 主要输入.....	479
11.19.2 主要工具与技术.....	434	12.4.2 主要工具与技术.....	480
11.19.3 主要输出.....	435	12.4.3 主要输出.....	481
11.20 实施定性风险分析.....	436	12.5 建设团队.....	481
11.20.1 主要输入.....	437	12.5.1 主要输入.....	483
11.20.2 主要工具与技术.....	437	12.5.2 主要输出.....	483
11.20.3 主要输出.....	439	12.6 管理团队.....	483
11.21 实施定量风险分析.....	439	12.6.1 主要输入.....	484
11.21.1 主要输入.....	440	12.6.2 主要工具与技术.....	485
11.21.2 主要工具与技术.....	441	12.7 管理沟通.....	486
11.21.3 主要输出.....	444		



12.7.1 主要输入.....	488
12.7.2 主要工具与技术.....	488
12.7.3 主要输出.....	489
12.8 实施风险应对.....	489
主要输入.....	490
12.9 实施采购.....	490
12.9.1 主要输入.....	492
12.9.2 主要输出.....	493
12.10 管理干系人参与.....	493
主要输入.....	494
12.11 本章练习.....	495
第 13 章 监控过程组.....	497
13.1 控制质量.....	498
13.1.1 主要输入.....	499
13.1.2 主要工具与技术.....	500
13.1.3 主要输出.....	502
13.2 确认范围.....	502
13.2.1 确认范围的关键内容.....	503
13.2.2 主要输入.....	504
13.2.3 主要输出.....	505
13.3 控制范围.....	505
13.3.1 主要输入.....	506
13.3.2 主要输出.....	506
13.4 控制进度.....	506
13.4.1 主要输入.....	508
13.4.2 主要工具与技术.....	508
13.4.3 主要输出.....	510
13.5 控制成本.....	510
13.5.1 主要输入.....	511
13.5.2 主要工具与技术.....	512
13.5.3 主要输出.....	514
13.6 控制资源.....	514
主要输入.....	515
13.7 监督沟通.....	516
主要输入.....	516
13.8 监督风险.....	517

13.8.1 主要输入.....	518
13.8.2 主要工具与技术.....	518
13.8.3 主要输出.....	519
13.9 控制采购.....	519
13.9.1 主要输入.....	521
13.9.2 主要工具与技术.....	522
13.9.3 主要输出.....	522
13.10 监督干系人参与.....	523
13.10.1 主要输入.....	524
13.10.2 主要工具与技术.....	524
13.10.3 主要输出.....	525
13.11 监控项目工作.....	525
13.11.1 主要输入.....	526
13.11.2 主要工具与技术.....	527
13.11.3 主要输出.....	528
13.12 实施整体变更控制.....	532
13.12.1 主要输入.....	533
13.12.2 主要输出.....	534
13.13 本章练习.....	534

第 14 章 收尾过程组..... 537

14.1 结束项目或阶段.....	538
14.1.1 主要输入.....	539
14.1.2 主要输出.....	540
14.2 收尾过程组的重点工作.....	541
14.2.1 项目验收.....	541
14.2.2 项目移交.....	542
14.2.3 项目总结.....	543
14.3 本章练习.....	544

第 15 章 组织保障..... 546

15.1 信息和文档管理.....	546
15.1.1 信息和文档.....	546
15.1.2 信息（文档）管理规则和方法.....	547
15.2 配置管理.....	549
15.2.1 基本概念.....	549
15.2.2 角色与职责.....	553

15.2.3 目标与方针.....	554	17.1.2 法律体系.....	580
15.2.4 管理活动.....	555	17.1.3 法的效力.....	582
15.3 变更管理.....	559	17.1.4 法律法规体系的效力层级.....	583
15.3.1 基本概念.....	559	17.1.5 信息系统集成项目管理中常用的 法律.....	584
15.3.2 角色与职责.....	561	17.2 标准规范.....	585
15.3.3 工作程序.....	562	17.2.1 标准和标准化.....	586
15.3.4 变更控制.....	564	17.2.2 标准分级与标准分类.....	587
15.3.5 版本发布和回退计划.....	566	17.2.3 我国标准的编号及名称.....	588
15.4 本章练习.....	566	17.2.4 我国标准的有效期.....	591
第 16 章 监理基础知识.....	569	17.2.5 信息系统集成项目管理中常用的 标准规范.....	591
16.1 监理的意义和作用.....	569	17.3 本章练习.....	593
16.2 监理相关概念.....	571	第 18 章 职业道德规范.....	595
16.3 监理依据.....	573	18.1 基本概念.....	595
16.4 监理内容.....	574	18.2 项目管理工程师职业道德规范.....	596
16.5 监理要素.....	575	18.3 项目管理工程师岗位职责.....	597
16.5.1 监理合同.....	575	18.4 项目管理工程师对项目团队的责任.....	597
16.5.2 监理服务能力.....	576	18.5 提升个人道德修养水平.....	598
16.6 本章练习.....	579	18.6 本章练习.....	599
第 17 章 法律法规和标准规范.....	580	参考文献.....	600
17.1 法律法规.....	580		
17.1.1 法与法律.....	580		

第3章 信息技术服务

信息技术服务（Information Technology Service，IT 服务）通常是指供方为需方提供如何开发、应用信息技术的服务，以及供方以信息技术为手段提供支持需方业务活动的服务。在新时代，IT 服务持续融合创新的环境下，IT 服务突破了单纯的供方为需方提供服务的限制，IT 服务的内涵定义为：组织为达成用户期望的结果，利用信息技术为用户交付价值的活动。

IT 服务是随着信息技术的发展和信息技术在各行业的深入应用而产生的一种新兴业态，是数字经济的重要内容之一，是现代信息服务业的重要组成部分。IT 服务是信息技术与服务的结合，既具有服务的特征，又具有信息技术的独特特征。

3.1 内涵与外延

随着网络的快速发展，包括互联网的泛化以及数据要素的驱使等，使其上的应用能够通过多种终端与个人紧密结合，创造和改变了众多组织及个人的应用习惯和业务模式等，为服务提供了新的实现手段，也赋予了服务更多的内涵，除软硬件技术支持服务、服务外包、IT 咨询、IT 培训等服务外，以新媒体、社交网络、数据开发等为代表的新领域开始蓬勃发展，IT 服务走向多元化发展模式，深度影响产业经济发展和社会治理改革等。

3.1.1 服务的特征

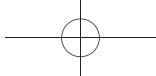
服务是一种通过提供必要的手段和方法，满足服务接受者需求的“过程”，其外延是指具备服务本质的一切服务，例如餐饮服务、零售服务、IT 服务等。服务的特征包括：无形性、不可分离性、可变性和不可储存性等。

（1）无形性。无形性指服务在很大程度上是抽象的和无形的。需方在购买之前一般无法看到、感觉到或触摸到，例如理发、听音乐会、到海边度假等。这一特性使得服务不容易向需方展示或沟通交流，因此需方难以评估其质量。

（2）不可分离性。不可分离性又叫同步性，指生产和消费是同时进行的，如照相、理发等。这一特性表明，需方只有参与到服务的生产过程中才能享受到服务。这一特性决定了服务质量管理对服务供方的重要性，其服务的态度和水平直接决定了需方对该项服务的满意度。因此，服务人员的筛选、培训和报酬标准等，对实现高标准的服务质量至关重要。

（3）可变性。可变性也叫异质性，指服务的质量水平会受到相当多因素的影响，因此会经常变化。服务以人中心，由于人与人的文化、修养、技术水平等存在差异，同一服务的品质会因操作者不同而不同；即使是同一操作者，由于时间、地点与心态的变化，服务质量也会随之变化。

（4）不可储存性。不可储存性也叫易逝性、易消失性，指服务无法被储藏起来以备将来使用、转售、延时体验或退货等。



3.1.2 IT服务的内涵

按照国务院发展研究中心的相关研究成果定义，信息服务包括信息传输服务、IT 服务和信息内容服务。在国家统计局最新修订的《国民经济行业分类》（GB/T 4754）中，已将 IT 服务划分为一个国民经济行业大类。在中国经济增长方式转变，产业结构优化升级，IT 应用不断深化，特别是在数字中国、数字经济、数字政府、数字化转型等的背景下，IT 服务新业态与技术创新不断涌现（如车联网、工业互联网、自媒体等），为 IT 服务业发展带来更多新的机遇和挑战。IT 服务除了具备服务的基本特征，还具备本质特征、形态特征、过程特征、阶段特征、效益特征、内部关联性特征、外部关联性特征等方面的内涵。

（1）本质特征。IT 服务的组成要素包括人员、过程、技术和资源。就 IT 服务而言，通常情况下是由具备匹配的知识、技能和经验的人员，合理运用资源，并通过规定过程向需方提供 IT 服务。

（2）形态特征。常见服务形态有 IT 咨询服务、设计与开发服务、信息系统集成服务、数据处理和运营服务、智能化服务及其他 IT 服务等。

（3）过程特征。IT 服务从项目级、组织级、量化管理级、数字化运营等逐步发展，是从计算机单机应用、网络应用、综合管理的逐步提升，具有连续不断和可持续发展的特征。

（4）阶段特征。由于 IT 的发展永无止境，信息基础设施和经济、市场环境的变迁，IT 服务也无终极目标。IT 服务是全方位的，无论需方还是供方都需要根据自身需要抓重点。分层次、分阶段地推进 IT 服务，提高 IT 的有效利用。

（5）效益特征。IT 服务的发展不同于以往对产品的技术改造，其效益的概念完全不同。通过对产品生产线的技术改造，提高质量、增加产能，其效果往往是单方面的，且容易显现。而 IT 服务的发展是服务系统进行深度开发和广泛利用，从整体上提高组织核心竞争力和管理水平，其效益是多方面的。

（6）内部关联性特征。IT 服务不仅依赖于技术创新，更依赖于业务模式创新。保持技术创新和业务模式创新的相互促进、有机融合，实现 IT 服务人才结构优化，建立 IT 服务管理规范，将从机制上为 IT 服务的发展创造条件。

（7）外部关联性特征。IT 服务依赖于国民经济和良性竞争的市场环境的形成，依赖于社会信息网络的不断进步，依赖于政府相应的政策支撑、配套人才的培养和产业链上下游组织 IT 应用的逐渐完善。

3.1.3 IT服务的外延

在各行业加速数字化转型的形势下，随着新一代信息技术的快速发展和应用，IT 服务的外延得到很大的拓展。新形势下，IT 服务包括基础服务、技术创新服务、数字化转型服务等。

（1）基础服务。基础服务指面向 IT 的基础类服务，主要包括咨询设计、开发服务、集成实施、运行维护、云服务和数据中心等。

（2）技术创新服务。技术创新服务指面向新技术加持下的新业态新模式的服务，主要包含智能化服务、数字服务、数字内容处理服务和区块链服务等。

（3）数字化转型服务。数字化转型服务指支撑和服务组织数字化服务开展和创新融合业务发展的服务，主要包括数字化转型成熟度推进服务、评估评价服务、数字化监测预警服务等。

(4) 业务融合服务。业务融合服务指信息技术及其服务与各行业融合的服务,如面向政务、广电、教育、应急、业财等行业。

3.1.4 IT服务业的特征

IT 服务业具有高知识和高技术含量、高集群性、服务过程的交互性、服务的非独立性、知识密集性、产业内部呈金字塔分布、法律和契约的强依赖性以及声誉机制等特征。

(1) 高知识和高技术含量。IT 服务业的提供者是生产过程中的专家组,多以技术资本、知识资本、人力资本为主要投入,产出中有密集的知识要素,因此 IT 服务业把日益专业化的技术、知识加入服务过程中,具有人力资源、技术、知识密集的特点。IT 服务业需要向需方转移高度专业化的知识,这是其区别于其他服务业的一个显著特征。

(2) 高集群性。IT 服务业在其空间上具有很高的集群性。IT 服务业的出现和发展都集中在大型中心城市。其中,中心城市具有及时准确的宏观政策、完善的基础设施、高智力的人力资源及发达的人力资源市场,这些因素为 IT 服务业发展提供了良好的条件。

(3) 服务过程的交互性。需方参与服务过程,IT 服务业不仅提供显性知识,还提供隐性知识,要实现隐性知识的传播需要通过专业人员与需方间进行大量的互动过程才能完成。

(4) 服务的非独立性。IT 服务业提供的是满足需方需求的解决方案,往往涉及多个领域的知识,许多 IT 服务业与高等院校、科研机构形成联盟,相互合作。因此除了自身具备的知识技术外,IT 服务业会将其他行业机构的技术与成果进行整合,这是 IT 服务业比较突出的特征。

(5) 知识密集性。IT 服务提供过程中的交互活动依赖个人的专业知识,因此,个人知识成为 IT 服务业的关键性资源,IT 服务业间的竞争更多是人才竞争。没有高素质人才,IT 服务业就成了无本之木。IT 服务业的从业人员需要具备完整的知识结构、丰富的专业知识和实践经验,方能满足需方的需求,帮助需方制定与实施完善的、适宜的解决方案。

(6) 产业内部呈金字塔分布。IT 服务产品差异性比较大,具有资金需求小、成本低、标准化程度不足等特点,因此进入壁垒相对较低。现代服务业内部结构呈金字塔分布,存在少数大型的组织和多数小型的组织。

(7) 法律和契约的强依赖性。IT 服务业在提供服务的供方与接受服务的需方间主要以签订服务协议或者契约的形式来确定相关服务事项,从而在双方间形成一种委托代理关系。因此,IT 服务业与法律和契约之间具有较强的依赖关系。

(8) 声誉机制。IT 服务业的生产和消费,由于空间和时间上的不可分性,使接受服务的需方事先无法观察到服务的质量,因此需方主要根据供方的声誉来确定对服务的支付意愿。反映供方声誉和质量的证明是决定选取和使用方面的重要因素,因此声誉机制对 IT 服务业的服务业务发展起着决定性作用。

3.2 原理与组成

2009 年在工业和信息化部软件服务业司的指导下,我国组建了 IT 服务标准工作组,全面推动国家 IT 服务标准建设,形成了 IT 服务标准库(Information Technology Service Standards,

ITSS)。ITSS 是一套体系化 IT 服务标准库，全面规范了 IT 服务产品及其组成要素，是我国 IT 服务最佳实践的总结和提升，也是我国从事 IT 服务开发、供应、推广和应用各类组织创新成果的固化。经过十多年的持续建设，ITSS 已经发布了包含国家标准、行业标准和社团标准等在内的百余项标准，是我国标准化和可信赖 IT 服务的主要指导依据。

3.2.1 IT服务原理

ITSS 给出了 IT 服务的基本原理，由能力要素、生存周期要素、管理要素组成，如图 3-1 所示。能力要素由人员（People）、过程（Process）、技术（Technology）和资源（Resource）组成，简称 PPTR。IT 服务生命周期由四个阶段组成，分别是战略规划（Strategy&planning）、设计实现（Design&implementation）、运营提升（Operation& promotion）、退役终止（Retirement &termination），简称为 SDOR。为了保证每个生命周期阶段及其过程的实施，需要通过监督提供保障，通过管理提供支撑。

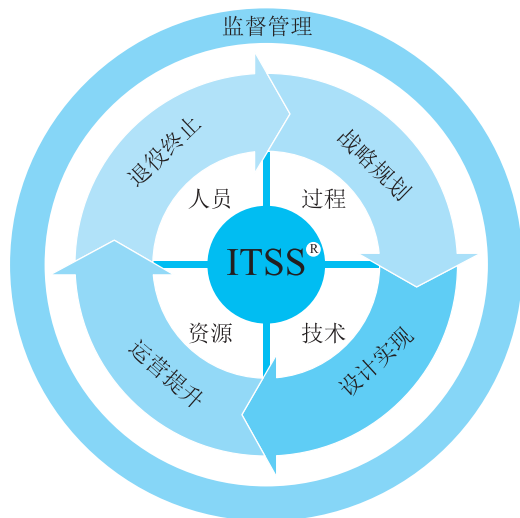


图 3-1 ITSS 基本原理

3.2.2 组成要素

ITSS 定义了 IT 服务由人员、过程、技术和资源组成，并对这些服务的组成要素进行标准化。另外，就 IT 服务而言，通常情况下是由具备匹配的知识、技能和经验的人员，合理运用资源，并通过规定的过程向需方提供服务。

1. 人员

人员是指 IT 服务生命周期中各类满足要求的人才的总称，ITSS 规定了提供 IT 服务的各类人员应具备的知识、经验和技能要求，目的是指导 IT 服务提供商根据岗位职责和管理要求“正确选人”。

一般而言，针对咨询设计、集成实施、运行维护和运营等典型的 IT 服务，所需要的人员包

括项目经理（如系统集成项目经理、服务项目经理）、系统分析师、架构设计师、系统集成工程师、信息安全工程师、系统评测工程师、服务工程师、服务定价师、客户经理和日常服务人员等，如图 3-2 所示。

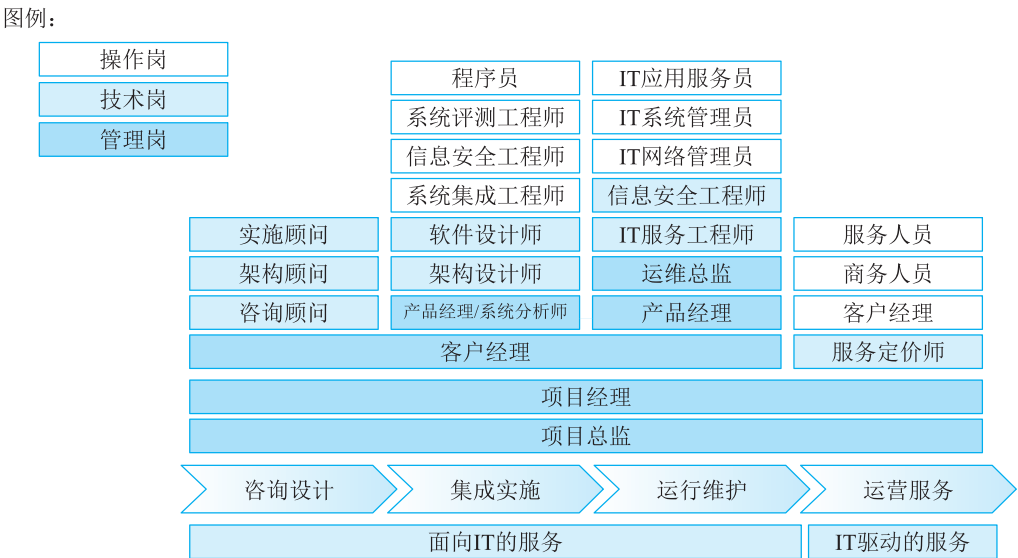


图 3-2 IT 服务所需要的人员

目前，针对 IT 服务人员，使得服务供方常面临如下挑战：

- 人员知识、技能和经验评估难；
- 不同人员交付同一服务的质量不一致；
- 人才流动率高，很难建设稳定的服务团队；
- 人才招聘难，很难形成合理的人力资源池；
- 人员专业化的必要性；
- 有助于建立与服务发展相适应的人才队伍，保障服务工作的连续性和稳定性；
- 有助于改进和完善人才培养模式，提高人才培养质量；
- 有助于优化人力资源管理，提高管理效率和降低管理成本。

2. 过程

过程是通过合理利用必要的资源，将输入转化为输出的一组相互关联和结构化的活动，是提高管理水平和确保服务质量的关键要素。ITSS 根据咨询设计、集成实施、运行维护等各种类型的 IT 服务，规定了应建立的过程和各个过程应实现的关键绩效指标（KPI），确保供方能“正确做事”。通过按照 ITSS 要求建立简洁、高效和协调的过程，能有效地将人员、技术和资源要素连接起来，指导服务人员按规定的方式方法正确地做事。

过程作为 IT 服务的核心要素之一，有明确的目标，可重复和可度量。各类 IT 服务的典型过程如图 3-3 所示。



图 3-3 各类 IT 服务的典型过程

对各类 IT 服务组织来说，过程要素所面临的挑战主要包括以下几个方面：

- 过程没有明确定义，完全按照操作人员的个人习惯执行；
- 过程定义不清晰，不具备按照过程管理思路执行的价值；
- 过程定义太复杂，执行效率严重下降甚至影响运营；
- 没有明确的过程目标，操作人员不清楚每一项活动应该做到什么；
- 对过程没有监督，不清楚过程的稳定性；
- 对过程没有考核，不能得到持续改进。

过程规范化的必要性主要包括以下几个方面：

- 确保过程可重复和可度量；
- 有效控制因未明确定义而引发的潜在风险；
- 通过对过程进行评价和度量，可持续提升过程的效率；
- 通过过程实现规范化管理，可持续提高服务质量；
- 通过规范化的过程管理，提高效率，减少人员和成本的投入。

3. 技术

技术是指交付满足质量要求的 IT 服务应使用的技术或应具备的技术能力，以及提供 IT 服务所必须的分析方法、架构和步骤。技术要素确保 IT 服务供方能“高效做事”，是提高 IT 服务质量方面重点考虑的要素，主要通过自有核心技术的研发和非自有核心技术的学习借鉴，持续提升提供服务过程中发现问题和解决问题的能力。

在提供服务过程中，可能面临各种问题、风险以及新技术和前沿技术应用所提出的新要求，供方应根据需方要求或技术发展趋势，具备发现和解决问题、风险控制、技术储备以及研发、应用新技术和前沿技术的能力。针对咨询设计、集成实施、运行维护等 IT 服务，常用的技术如图 3-4 所示。

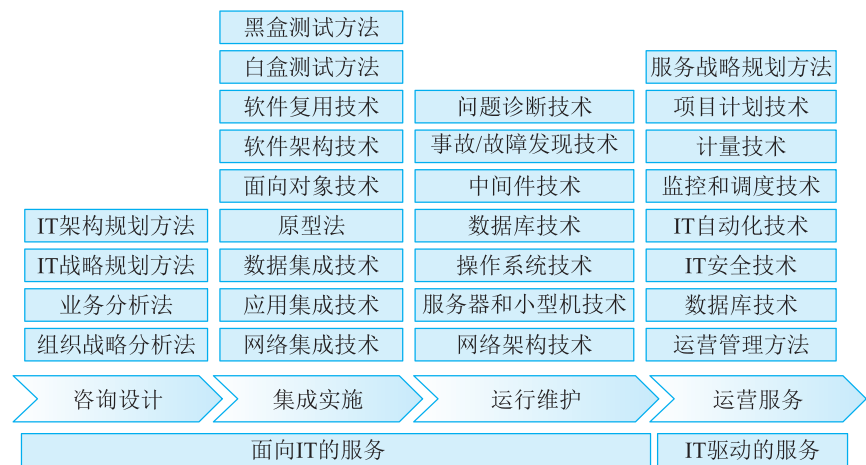


图 3-4 IT 服务常用的技术

- 对服务供需双方来说，技术要素所面临的挑战主要包括：
- 为满足组织的目标和需求，组织对IT技术的依赖程度越来越高；
 - 激烈的市场竞争使得组织对技术的要求越来越高；
 - 低成本、高效率的服务需求，对组织的技术研发和使用能力提出了更高的要求。

技术体系化的必要性主要体现在以下几个方面：

- 提高服务质量，降低服务成本；
- 减少人员流失带来的损失；
- 及时应用和推广成熟技术；
- 做好新技术的研发和储备；
- 在提供服务中使用一致的技术标准。

4. 资源

资源是指提供 IT 服务所依存和产生的有形及无形资产，如咨询服务供方为满足需方的需求，提供咨询服务所必须具备的知识、经验和工具等。资源要素确保供方能“保障做事”，主要由人员、过程和技术要素中被固化的成果和能力转化而成，同时又对人员、过程和技术要素提供有力的支撑和保障。

根据所提供的 IT 服务类型的不同，所需要的资源也不尽相同，但可以对其进行汇总。例如，咨询设计服务和 IT 运维服务所使用的资源包括知识库、工具库、专家库、备件库和服务台。常见的 IT 服务资源如图 3-5 所示。

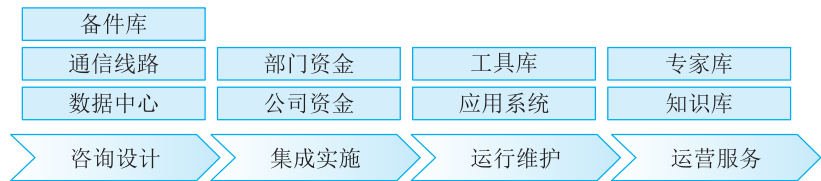


图 3-5 常见的 IT 服务资源

对于各类服务组织来说，资源要素所面临的挑战主要包括：

- 忽略资源的价值，投入不够，导致资源不足；
- 对资源的使用不重视，重复投资现象严重；
- 缺乏利用资源的统一规划，资源的利用率不高；
- 资源的更新不及时，与需求、技术研发脱节。

资源系统化的必要性主要体现在以下几个方面：

- 统筹资源开发利用，确保与运营、技术研发协调一致；
- 确保提供满足质量和成本要求的服务；
- 明确各类资源管理的要点，提高资源使用率；
- 结合服务发展需求，确保能及时更新资源，提高资源的使用率和使用质量。

3.3 服务生命周期

IT 服务生存周期是指 IT 服务从战略规划、设计实现、运营提升到退役终止的演变，如图 3-6 所示。IT 服务生命周期的引入，改变了 IT 服务在不同阶段相互割裂、独立实施的局面。同时，通过连贯的逻辑体系，以战略规划为指引，设计实现为准绳，通过服务运营实现价值转化，直至服务的退役终止。同时伴随着监督管理的不断完善，将服务中的不同阶段的不同过程有机整合为一个井然有序、良性循环的整体，使服务质量得以不断改进和提升。

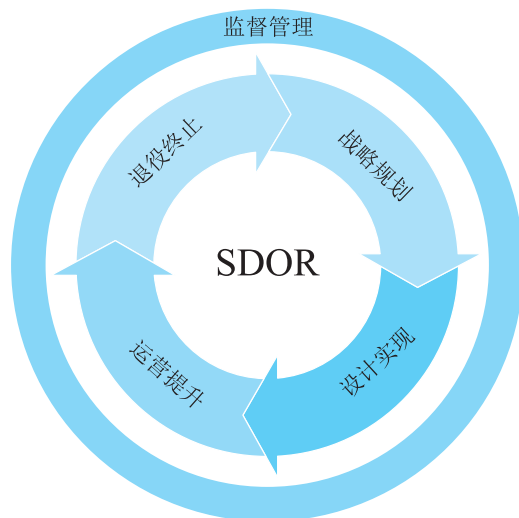
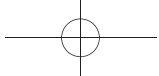


图 3-6 ITSS 定义的 IT 服务生命周期

（1）战略规划。战略规划是指从组织战略出发，以需求为中心，参照 ITSS 对 IT 服务进行战略规划，为 IT 服务的设计实现做好准备，以确保提供满足供需双方需求的 IT 服务。

（2）设计实现。设计实现是指依据战略规划，定义 IT 服务的体系结构、组成要素、要素特征以及要素之间的关联关系，建立管理体系、部署专用工具以及服务解决方案。



第 6 章 数据工程

数据工程是信息系统的基础工程。数据工程围绕数据的生命周期及管理要求，研究数据从采集清洗到应用服务的全过程，为信息系统运行提供可靠的数据基础，为信息系统之间的数据共享提供安全、高效的保障，为信息系统实现互连、互通、互操作提供支撑。组织的数据工程相关能力是其建设数据要素的关键，是组织数据资源化、数据标准化、数据资产化、数据价值化的重要手段。

6.1 数据采集和预处理

有效且高质量的数据获取是组织数据要素建设的重要活动，关系到组织数据的质量基础、容量规模、价值化开发等。广泛多元的数据采集以及必要的预处理，是支撑和保障数据获取的主要活动。

6.1.1 数据采集

数据采集又称数据收集，是指根据用户需要收集相关数据的过程。采集的数据类型包括结构化数据、半结构化数据、非结构化数据。结构化数据是以关系型数据库表管理的数据；半结构化数据是指非关系模型的、有基本固定结构模式的数据，例如日志文件、XML 文档、E-mail 等；非结构化数据是指没有固定模式的数据，如所有格式的办公文档、文本、图片、HTML、各类报表、图像和音频 / 视频信息等。

数据采集的方法可分为传感器采集、系统日志采集、网络采集和其他数据采集等。

传感器采集是通过传感器感知相应的信息，并将这些信息按一定规律变换成电信号或其他所需的信息输出，从而获取相关数据，是目前应用非常广泛的一种采集方式。数据采集传感器包括重力感应传感器、加速度传感器、光敏传感器、热敏传感器、声敏传感器、气敏传感器、流体传感器、放射线敏感传感器、味敏传感器等。

系统日志采集是通过平台系统读取、收集日志文件变化。系统日志记录系统中硬件、软件和系统运行情况及问题的信息。系统日志一般为流式数据，数据量非常庞大，常用的采集工具有 Logstash、Filebeat、Flume、Fluentd、Logagent、rsyslog、syslog-ng 等。

网络采集是指通过互联网公开采集接口或者网络爬虫等方式从互联网或特定网络上获取大量数据信息的方式，是实现互联网数据或特定网络采集的主要方式。数据采集接口一般通过应用程序接口（API）的方式进行采集。网络爬虫（Web Crawler/Web Spider）是根据一定的规则来提取所需要信息的程序。根据系统结构和实现技术，网络爬虫可分为通用网络爬虫（General Purpose Web Crawler）、聚焦网络爬虫（Focused Web Crawler）、增量式网络爬虫（Incremental Web Crawler）、深层网络爬虫（Deep Web Crawler）等类型。

除此之外，还有一些其他的数据采集方式，如通过与数据服务商合作，使用特定数据采集方式获取数据。

6.1.2 数据预处理

数据的预处理一般采用数据清洗的方法来实现。数据预处理是一个去除数据集重复记录，发现并纠正数据错误，并将数据转换成符合标准的过程，从而使数据实现准确性、完整性、一致性、唯一性、适时性、有效性等。一般来说，数据预处理主要包括数据分析、数据检测和数据修正 3 个步骤，如图 6-1 所示。

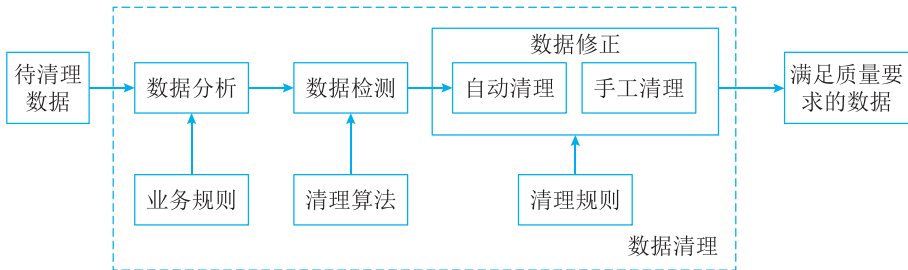


图 6-1 数据预处理的流程

- (1) 数据分析：是指从数据中发现控制数据的一般规则，比如字段域、业务规则等。通过对数据的分析，定义出数据清理的规则，并选择合适的算法。
- (2) 数据检测：是指根据预定义的清理规则及相关数据清理算法，检测数据是否正确，比如是否满足字段域、业务规则等，或检测记录是否重复。
- (3) 数据修正：是指手工或自动地修正检测到的错误数据或重复的记录等。

6.1.3 数据预处理方法

一般而言，需要进行预处理的数据主要包括数据缺失、数据异常、数据不一致、数据重复、数据格式不符等情况，针对不同问题需要采用不同的数据处理方法。

1. 缺失数据的预处理

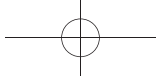
数据缺失产生的原因主要分为环境原因和人为原因，需要针对不同的原因采取不同的数据预处理方法，常见的方法有删除缺失值、均值填补法、热卡填补法等。

删除缺失值是最常见的、简单有效的方法，当样本数很多的时候，并且出现缺失值的样本占整个样本的比例相对较小时，可以将有缺失值的样本直接丢弃。

均值填补法是根据缺失值的属性相关系数最大的那个属性把数据分成几个组，再分别计算每个组的均值，用均值代替缺失数值。

热卡填补法通过在数据库中找到一个与包含缺失值变量最相似的对象，然后采用相似对象的值进行数据填充。

缺失数据预处理的其他方法还有最近距离决定填补法、回归填补法、多重填补法、K-最近邻法、有序最近邻法、基于贝叶斯的方法等。



2. 异常数据的预处理

对于异常数据或有噪声的数据，如超过明确取值范围的数据、离群点数据，可以采用分箱法和回归法来进行处理。

分箱法通过考察数据的“近邻”（即周围的值）来平滑处理有序的数据值，这些有序的值被分布到一些“桶”或“箱”中，进行局部光滑。一般而言，宽度越大，数据预处理的效果越好。

回归法用一个函数拟合数据来光滑数据，消除噪声。线性回归涉及找出拟合两个属性（或变量）的“最佳”直线，使得一个属性能够预测另一个。多线性回归是线性回归的扩展，它涉及多于两个属性，并且数据拟合到一个多维面。

3. 不一致数据的预处理

不一致数据是指具有逻辑错误或者数据类型不一致的数据，如年龄与生日数据不符。这一类数据的清洗可以使用人工修改，也可以借助工具来找到违反限制的数据，如知道数据的函数依赖关系，可以通过函数关系修改属性值。但是大部分的不一致情况都需要进行数据变换，即定义一系列的变换纠正数据，有一些商业工具可以提供数据变换的功能，例如数据迁移工具和 ETL 工具等。

4. 重复数据的预处理

数据本身存在的或数据清洗后可能会产生的重复值。重复值的存在会影响后续模型训练的质量，造成计算及存储浪费。去除重复值的操作一般最后进行，可以使用 Excel、VBA（Visual Basic 宏语言）、Python 等工具处理。

5. 格式不符数据的预处理

一般人工收集或者应用系统用户填写的数据，容易存在格式问题。一般需要将不同类型的数据内容清洗成统一类型的文件和统一格式，如将 TXT、CSV、Excel、HTML 以及 PDF 清洗成统一的 Excel 文件，将显示不一致的时间、日期、数值或者内容中有空格、单引号、双引号等情况进行格式的统一调整。

6.2 数据存储及管理

通过数据采集和预处理获得的数据，往往是组织具备较高价值的数字资源，确保这些数据得到适当的保管和管理，是数据价值化的基础，组织往往根据数据规模和数据的重要性等，采用最合适的存储介质、存储方法、管理体系、管理措施等。

6.2.1 数据存储

数据存储就是根据不同的应用环境，通过采取合理、安全、有效的方式将数据保存到物理介质上，并能保证对数据实施有效的访问。其中包含两个方面：一是数据临时或长期驻留的物理媒介；二是保证数据完整、安全存放和访问而采取的方式或行为。数据存储就是把这两个方面结合起来，提供完整的解决方案。

1. 数据存储介质

数据存储首先要解决的是存储介质的问题。存储介质是数据存储的载体，是数据存储的基础。存储介质并不是越贵越好、越先进越好，要根据不同的应用环境，合理选择存储介质。存储介质的类型主要有磁带、光盘、磁盘、内存、闪存、云存储等，其描述如表 6-1 所示。

表 6-1 常见数据存储介质的描述

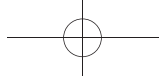
介质	描述
磁带	磁带是存储成本低、容量大的存储介质，主要包括磁带机、自动加载磁带机和磁带库。其主要的缺点就是速度比较慢
光盘	光盘的全称是高密度盘（Compact Disk），常见的格式有 VCD（Video Compact Disk）和 DVD（Digital Video Disk）两种，前者能提供 700MB 左右的空间，后者容量要大得多，可提供 4.7GB ~ 60GB 的存储空间。光盘具有 3 个显著特点：一是光盘上的数据具有只读性；二是不受电磁的影响；三是光盘容易大量复制。这些特点使得光盘特别适用用来对数据进行永久性归档备份
磁盘	利用磁盘存储数据时，一般采用独立冗余磁盘阵列 RAID（Redundant Array of Independent Disks）。RAID 将数个单独的磁盘以不同的组合方式形成一个逻辑磁盘，不仅提高了磁盘读取的性能，也增强了数据的安全性
内存	内存是计算机用于存放 CPU 中的运算数据，与硬盘等外部存储器交换数据的硬件。内存的性能决定了计算机运行的稳定性、反应速率。通常来说，内存数据会在断电后丢失所有数据
闪存	闪存是一种固态技术，使用闪存芯片来写入和存储数据，具有集内存的访问速度和存储持久性于一体的特点，常作为磁盘的替代品
云存储	与将数据存储到本地硬盘驱动器或存储网络相比，云存储提供了一种可扩展的替代方案，将数据存储在异地位置，可通过公共互联网或者专用私有网络进行访问

2. 存储形式

一般而言，主要有 3 种形式来记录和存储数据，分别是文件存储、块存储和对象存储，如表 6-2 所示。

表 6-2 主要数据存储形式的描述

存储形式	描述
文件存储	文件存储也称为文件级或基于文件的存储，是一种用于组织和存储数据的分层存储方法。换言之，数据存储在文件中，文件被组织在文件夹中，文件夹则被组织在目录和子目录的层次结构下
块存储	块存储有时也称为块级存储，是一种用于将数据存储成块的技术。这些块随后作为单独的部分存储，每个部分都有唯一的标识符。对于需要快速、高效和可靠地进行数据传输的计算场景，开发人员一般倾向于使用块存储
对象存储	对象存储通常称为基于对象的存储，是一种用于处理大量非结构化数据的数据存储架构。这些数据无法轻易组织到具有行和列的传统关系数据库中，或不符合其要求，如电子邮件、视频、照片、网页、音频文件、传感器数据以及其他类型的媒体和 Web 内容（文本或非文本）



3. 存储管理

存储管理在存储系统中的地位越来越重要，例如，如何提高存储系统的访问性能，如何满足数据量不断增长的需要，如何有效地保护数据、提高数据的可用性，如何满足存储空间的共享等。存储管理的具体内容如表 6-3 所示。

表 6-3 存储管理的主要内容

管理方面	主要内容
资源调度管理	资源调度管理的功能主要是添加或删除存储节点，编辑存储节点的信息，设定某类型存储资源属于某个节点，或者设定这些资源比较均衡地存储到节点上。它包含存储控制、拓扑配置以及各种网络设备（如集线器、交换机、路由器和网桥等）的故障隔离
存储资源管理	存储资源管理是一类应用程序，它们管理和监控物理和逻辑层次上的存储资源，从而简化资源管理，提高数据的可用性。被管理的资源主要是存储硬件，如 RAID、磁带以及光盘库。存储资源管理不仅包括监控存储系统的状况、可用性、性能以及配置情况，还包括容量和配置管理以及事件报警等，从而提供优化策略
负载均衡管理	负载均衡是为了避免存储资源由于资源类型、服务器访问频率和时间不均衡造成浪费或形成系统瓶颈而平衡负载的技术
安全管理	存储系统的安全主要是防止恶意用户攻击系统或窃取数据。系统攻击大致分为两类：一类以扰乱服务器正常工作为目的，如拒绝服务（DoS）攻击、勒索病毒攻击等；另一类以入侵或破坏服务器为目的，如窃取数据、修改网页等

6.2.2 数据归档

因数据量海量增长和存储空间容量有限的矛盾，需要制定合理的数据归档方案，并及时清除过时的、不必要的数据，从而保证数据库性能的稳定。

数据归档是将不活跃的“冷”数据从可立即访问的存储介质迁移到查询性能较低、低成本、大容量的存储介质中，这一过程是可逆的，即归档的数据可以恢复到原存储介质中。数据归档策略需要与业务策略、分区策略保持一致，以确保最需要数据的可用性和系统的高性能。在开展数据归档活动时，有以下 3 点值得注意：

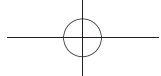
（1）数据归档一般只在业务低峰期执行。因为数据归档过程需要不断地读写生产数据库，这个过程将会大量使用网络，会对线上业务造成压力。

（2）数据归档之后，将会删除生产数据库的数据，将会造成数据空洞，即表空间并未及时释放，若长时间没有新的数据填充，会造成空间浪费的情况。

（3）如果数据归档影响了线上业务，一定要及时止损，结束数据归档，进行问题复盘，及时找到问题和解决方案。

6.2.3 数据备份

数据备份是为了防止由于各类操作失误、系统故障等意外原因导致的数据丢失，而将整个应用系统的数据或一部分关键数据复制到其他存储介质上的过程。这样做是为了保证当应用系统的数据不可用时，可以利用备份的数据进行恢复，尽量减少损失。



（1）完全备份（Full Backup）：每次都对需要进行备份的数据进行全备份。当数据丢失时，用完全备份下来的数据进行恢复即可。这种备份主要有两个缺点：一是由于每次都对数据进行全备份，会占用较多的服务器、网络等资源；二是在备份数据中有大量的数据是重复的，对备份介质资源的消耗往往也较大。

（2）差分备份（Differential Backup）：每次所备份的数据只是相对上一次完全备份之后发生变化的数据。与完全备份相比，差分备份所需时间短，而且节省了存储空间。另外，差分备份的数据恢复很方便，管理员只需两份备份数据，如星期日的完全备份数据和故障发生前一天的差分备份数据，就能对系统数据进行恢复。

（3）增量备份（Incremental Backup）：每次所备份的数据只是相对于上一次备份后改变的数据。这种备份策略没有重复的备份数据，节省了备份数据存储空间，缩短了备份的时间，但是当进行数据恢复时就会比较复杂。如果其中有一个增量备份数据出现问题，那么后面的数据也就无法恢复了。因此增量备份的可靠性没有完全备份和差分备份高。

6.2.4 数据容灾

数据备份是数据容灾的基础。传统的数据备份主要采用磁带进行冷备份，备份磁带一般存放在机房中进行统一管理，一旦整个机房出现灾难，如火灾、盗窃和地震等时，这些备份磁带也随之毁灭，起不到任何容灾作用。

因此，真正的数据容灾就是要避免传统冷备份的先天不足，它在灾难发生时能全面、及时地恢复整个系统。容灾按其灾难恢复能力的高低可分为多个等级，例如，国际标准 SHARE 78 定义的容灾系统有 7 个等级，从最简单的仅在本地进行磁带备份，到将备份的磁带存储在异地，再到建立应用系统实时切换的异地备份系统。恢复时间也可以从几天到小时级到分钟级、秒级或零数据丢失等。从技术上看，衡量容灾系统有两个主要指标，即 RPO（Recovery Point Object，恢复点目标）和 RTO（Recovery Time Object，恢复时间目标），其中 RPO 代表了当灾难发生时允许丢失的数据量，而 RTO 则代表了系统恢复的时间。

数据容灾的关键技术主要包括远程镜像技术和快照技术。

（1）远程镜像技术。远程镜像技术是在主数据中心和备份中心之间进行数据备份时用到的远程复制技术。镜像是在两个或多个磁盘子系统上产生同一个数据镜像视图的数据存储过程，一个称为主镜像；另一个称为从镜像。按主从镜像所处的位置分为本地镜像和远程镜像。本地镜像的主从镜像处于同一个 RAID 中，而远程镜像的主从镜像通常分布在城域网或广域网中。由于远程镜像在远程维护数据的镜像，因此在灾难发生时，存储在异地的数据不会受到影响。

（2）快照技术。所谓快照，就是关于指定数据集的一个完全可用的复制，该复制是相应数据在某个时间点（复制开始的时间点）的映像。快照的作用有两个：①能够进行在线数据恢复，可以将数据恢复成快照产生时间点时的状态；②为用户提供另外一个数据访问通道，比如在原数据在线运行时，利用快照数据进行其他系统的测试、应用开发验证、数据分析、数据模型训练等。

6.3 数据治理和建模

数据治理是开展数据价值化活动的基础，关注对数字要素的管控能力，覆盖组织对数据相关活动的统筹、评估、指导和监督等工作，需要重点关注元数据、数据标准化、数据质量、数据模型和数据建模等方面的内容。

6.3.1 元数据

简单来说，元数据是关于数据的数据（Data About Data）。在信息技术及其服务行业，元数据往往被定义为提供关于信息资源或数据的一种结构化数据，是对信息资源的结构化描述。其实质是用于描述信息资源或数据的内容、覆盖范围、质量、管理方式、数据的所有者、数据的提供方式等有关的信息。

1. 信息对象

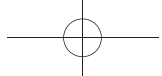
元数据描述的对象可以是单一的全文、目录、图像、数值型数据以及多媒体（声音、动态图像）等，也可以是多个单一数据资源组成的资源集合，或是这些资源的生产、加工、使用、管理、技术处理、保存等过程及其过程中产生的参数的描述等。

2. 元数据体系

根据信息对象从产生到服务的生命周期中，元数据描述和管理内容的不同以及元数据作用的不同，可以将元数据分为多种类型，从最基本的资源内容描述元数据开始，到指导描述元数据的元元数据，形成了一个层次分明、结构开放的元数据体系，如图 6-3 所示。

信息内容	内容元数据
	标记数字对象内容及结构的元数据
内容对象	专门元数据
	描述单一数字对象的内容、属性及外在特征的元数据
内容对象集合	资源集合元数据
	按照科学、主题、资源类型、用户范围、生成过程、使用管理范围形成的信息资源集合的描述
对象的管理与保存	管理元数据
	数字对象加工、存档、结构、技术处理、存取、控制、版权管理以及相关系统等方面的信息描述
对象的服务 服务过程 服务系统	服务元数据
	数字资源服务的揭示与表现、服务过程、服务系统等方面的相关信息的描述
元数据的 管理	元元数据
	对元数据的标记语言、格式语言、标识符、扩展机制、转换机制等的描述

图 6-3 元数据体系与元数据类型



元数据为数据的管理、发现和获取提供了一种实际而简便的方法。通过元数据，数据的使用者能够对数据进行详细、深入的了解，包括数据的格式、质量、处理方法和获取方法等各方面细节，可以利用元数据进行数据维护、历史资料维护等，具体作用包括描述、资源发现、组织管理数据资源、互操作性、归档和保存数据资源等，如表 6-5 所示。

表 6-5 元数据对数据使用者的作用说明

作用	说明
描述	用于描述数据的内容、覆盖范围、质量、管理方式、数据的所有者、数据的提供方式等信息，是数据与用户之间的桥梁
资源发现	元数据起到如同好的目录一样的作用，帮助用户便捷、快速地检索和确认所需要的资源
组织管理数据资源	随着基于 Web 的信息资源指数级的增长，按照用户和主题来链接信息资源的复合式网站或门户网站的作用越来越大，但是这种链接是用名称或位置硬编码在 HTML 文件中的，是一些静态的 Web 网页。利用存储在数据库中的元数据来组织动态网页，将更有效率和普遍意义，可以用软件工具为 Web 应用程序自动析取信息资源
互操作性	互操作性是不同硬件、软件平台、数据结构和接口以最少的内容和功能损失进行数据交换的能力。利用元数据描述的数据资源既可以被人类也可以被机器理解。使用良好定义的元数据语义和共享的转换协议，分布在网络上的数据资源就可以更加容易地被查找和转换，从而提高系统之间的互操作性
归档和保存数据资源	数据是脆弱的，它可能被无意识或有意识地破坏、修改，也可能由于存储介质、硬件和软件技术的变化而使得它不能使用。元数据是使数据资源保存下来，并能在将来继续被访问的关键，归档和保存数据需要特殊的数据元素来跟踪数字对象的来源（它来自何处、保护条件、保存责任等）。数据元素详细描述数字对象的物理特性和适应未来技术变化的行为

6.3.2 数据标准化

数据标准化主要为复杂的信息表达、分类和定位建立相应的原则和规范，使其简单化、结构化和标准化，从而实现信息的可理解、可比较和可共享，为信息在异构系统之间实现语义互操作提供基础支撑。数据标准化的主要内容包括元数据标准化、数据元标准化、数据模式标准化和数据分类与编码标准化。

在数据标准化活动中，要依据信息需求，并参照现行数据标准、信息系统的运行环境以及法规、政策和指导原则，在数据管理机构、专家组和开发者共同参与下，运用数据管理工具，得到注册的数据元素、物理模式和扩充的数据模型。数据标准化阶段的具体过程包括确定数据需求、制定数据标准、批准数据标准和实施数据标准。

（1）确定数据需求。本阶段将产生数据需求及相关的元数据、域值等文件。在确定数据需求时应考虑现行法规、政策，以及现行的数据标准。

（2）制定数据标准。本阶段要处理“确定数据需求”阶段提出的数据需求。如果现有的数据标准不能满足该数据需求，可以建议制定新的数据标准，也可建议修改或者封存已有的数据

标准。推荐的、新的或修改的数据标准记录在数据字典中。这个阶段将产生供审查和批准的成套建议。

(3) 批准数据标准。本阶段的数据管理机构对提交的数据标准建议、现行数据标准的修改或封存建议进行审查。一经批准，该数据标准将扩充或修改数据模型。

(4) 实施数据标准。本阶段涉及在各信息系统中实施和改进已批准的数据标准。

6.3.3 数据质量

数据质量指在特定的业务环境下，数据满足业务运行、管理与决策的程度，是保证数据应用效果的基础。数据质量管理是指运用相关技术来衡量、提高和确保数据质量的规划、实施与控制等一系列活动。衡量数据质量的指标体系包括完整性、规范性、一致性、准确性、唯一性、及时性等。数据质量是一个广义的概念，是数据产品满足指标、状态和要求能力的特征总和。

(1) 数据质量描述。数据质量可以通过数据质量元素来描述，数据质量元素分为数据质量定量元素和数据质量非定量元素。

(2) 数据质量评价过程。数据质量评价过程是产生和报告数据质量结果的一系列步骤，如图 6-4 所示描述了数据质量评价过程。

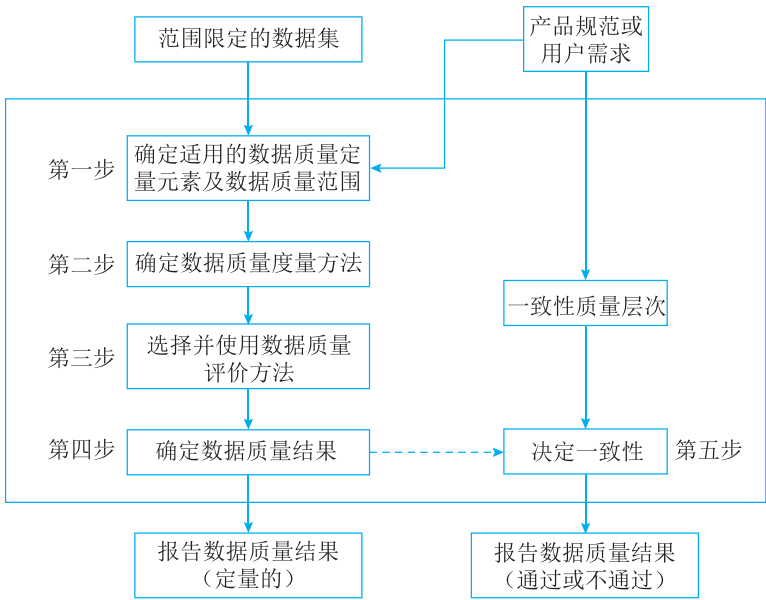


图 6-4 数据质量评价过程

(3) 数据质量评价方法。数据质量评价程序是通过应用一个或多个数据质量评价方法来完成的。数据质量评价方法分为直接评价法和间接评价法。直接评价法通过将数据与内部或外部的参照信息（如理论值等）进行对比来确定数据质量，间接评价法利用数据相关信息（如对数据源、采集方法等的描述）推断或评估数据质量。

(4) 数据质量控制。数据产品的质量控制在前期控制和后期控制两大部分。前期控制包