数据科学与大数据技术系列

SQL Server 2017 数据库分析处理技术

张延松 ◎ 编著

中国工信出版集团

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

数据科学与大数据技术系列

SQL Server 2017 数据库分析处理技术

张延松　编著

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

内 容 简 介

本书内容主要分为三部分：第 1 部分导论，介绍 SQL Server 2017 的安装及配置方法、数据导入方法和工具，并且通过数据可视化技术介绍数据分析处理技术的基本需求、数据模型及实现方法；第 2 部分数据库基础知识与 SQL 实践，介绍关系数据库基础理论、数据库基础实现技术、SQL 命令及查询实现技术、数据库实现新技术等相关知识；第 3 部分数据仓库和 OLAP 基础，介绍数据仓库的基本概念及相关理论、OLAP 的基本概念及相关操作、基于企业 Benchmark 的 OLAP 实践案例。

本书采用面向数据完整生命周期的贯穿式案例教学方法，以数据的采集、加载、管理、处理、分析、优化、数据可视化、多维展示、数据挖掘等从起点到终点的案例式处理过程，介绍数据分析处理全生命周期中相关的技术，使读者掌握全面的数据库分析处理技术，增强读者独立解决实际问题的能力。

本书既可以作为普通高等学校数据库课程教材，也可以作为数据库系统实现技术、数据仓库与数据分析技术等研究生课程的先行课教材。

图书在版编目（CIP）数据

SQL Server 2017 数据库分析处理技术 / 张延松编著. —北京：电子工业出版社，2019.8
（数据科学与大数据技术系列）
ISBN 978-7-121-37278-0

Ⅰ. ①S… Ⅱ. ①张… Ⅲ. ①关系数据库系统－高等学校－教材 Ⅳ. ①TP311.138

中国版本图书馆 CIP 数据核字（2019）第 179259 号

策划编辑：石会敏
责任编辑：底 波
印 刷：北京捷迅佳彩印刷有限公司
装 订：北京捷迅佳彩印刷有限公司
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：787×1 092 1/16 印张：23.5 字数：601.6 千字
版 次：2019 年 8 月第 1 版
印 次：2019 年 8 月第 1 次印刷
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）88254888，88258888。

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。

本书咨询联系方式：（010）88254537。

张延松

1973年出生，博士，副教授。

2010年于中国人民大学信息学院获得工学博士学位，2010—2012年在中国人民大学中国调查与数据中心从事博士后研究工作，2012年至今任职于中国人民大学信息学院，主要从事高性能数据库、内存数据库、数据仓库、OLAP、新硬件数据库技术等领域的研究工作，现开设数据库应用技术、大数据分析计算机基础、内存数据库等课程。

自2010年以来，以主要参与人或主持人身份参加了工信部核高基重大专项课题、国家863计划项目、国家自然科学基金重点及面上项目、北京市自然基金项目、华为校企合作项目等多项内存数据库及新硬件数据库技术方向的课题研究， 在国内外主要学术会议和期刊发表二十余篇学术论文，已申请二十余项国内外专利，并已获得4项美国专利、14项国内专利授权，出版学术专著2部，出版教材3部。作为参与人获得中国计算机学会2015年度科技进步奖一等奖，2016年教育部科技进步奖一等奖，2017年北京第十四届哲学社会科学优秀成果奖二等奖，2018年度国家科学技术进步奖二等奖。

前　言

数据是信息社会时代最重要的资源，数据库是面向数据管理的技术，不仅是计算机系统中重要的系统软件之一，而且也是以数据为中心的信息社会的基础支撑技术。在大数据时代，通过对海量数据的分析获得企业及社会发展所需要的有价值的信息是，深入挖掘大数据价值的重要手段，数据库技术也从面向电信、金融、航空订票、企业交易、电子商务等传统领域的事务处理扩展到对企业级海量数据的分析处理，并且结合大数据处理技术、新硬件技术等进一步升级了数据库大数据分析处理能力。传统的数据库以结构化数据管理为主，随着大数据多样化数据管理需求的增长，当前主流数据库提供了非结构化数据管理能力，如对 XML、JSON 等半结构化数据管理的支持及对 Hadoop 等大数据管理平台的支持，通过与数据可视化、数据分析处理等主流的大数据分析处理技术相融合的能力，数据库成为大数据时代一个重要的基础数据管理与分析处理平台。

当前，数据库技术的发展趋势呈现多个显著特点：新型内存查询型处理引擎集成到传统的磁盘处理引擎中，以提升数据库性能；从传统的行存储引擎转换为列存储引擎，以提高大数据分析处理性能；支持 XML、JSON 等非结构化数据管理；将 R、Python 等机器学习和分析软件集成到数据库查询引擎，以支持数据库中扩展的分析处理能力；增加与 Hadoop 等大数据管理平台集成工具，以支持数据库对 Hadoop 大数据平台的访问支持。随着数据库技术的升级与变革，传统数据库学科的理论知识与实践技能也需要进一步扩展，以更好地适应大数据时代数据库的理论体系和实现技术。

近年来，随着大数据分析处理需求的不断增长，企业级数据分析处理越来越多地成为数据库应用的主要任务之一，这需要更多具有不同学科知识背景的数据分析人员直接面对企业级数据平台，并掌握足够的数据库知识来完成企业级数据分析处理任务，这种应用需求要求降低数据库的技术门槛，以方便非计算机专业背景的数据分析人员使用数据库平台进行数据分析，同时也需要向非计算机专业人员系统地介绍数据库分析处理的完整技术框架。本书以 SQL Server 2017 平台为基础，以数据库分析处理技术为主线，以案例教学方式，系统地介绍数据库的基本理论、SQL 操作实践、数据库实现技术基本原理、数据仓库基本理论、OLAP 基本概念与实现技术、数据挖掘、数据可视化技术等，使读者通过实际的案例掌握以数据库为基础的数据分析处理技术所涉及的关系模型、存储模型、查询处理模型、查询优化模型、多维分析处理模型、OLAP 模型、数据挖掘模型、数据可视化模型等相关理论与实践技术，实现从数据到分析的完

整处理过程，较全面地理解现代数据库软件的体系结构和实现技术。

本书内容主要分为三部分。

1. 导论 介绍 SQL Server 2017 的安装及配置方法，数据导入方法和工具，并且通过数据可视化技术介绍数据分析处理技术的基本需求、数据模型及实现方法。

2. 数据库基础知识与 SQL 实践 介绍关系数据库基础理论、数据库基础实现技术、SQL 命令及查询实现技术、数据库实现新技术等相关知识。

3. 数据仓库和 OLAP 基础 介绍数据仓库的基本概念及相关理论、OLAP 的基本概念及相关操作、基于企业 Benchmark 的 OLAP 实践案例。

本书面向数据库分析处理技术，将通常分散于数据库基础、数据库实现技术、数据仓库与 OLAP、数据挖掘、商业智能与数据可视化技术等相关教材中的内容以统一的案例形式贯穿于本书始终，使读者能够通过案例实践，全面掌握数据分析处理的完整过程，从底层的数据库平台的数据组织与管理，到顶层面向分析模型的数据分析处理及可视化数据展示，实现理论与实践的统一。鉴于软件中代码的大小写对程序执行结果不会有影响，为体现其真实性，故本书里的截图保留了作者提供的原图，未进行大小写统一处理。

本书采用深入浅出的方法，结合案例教学模式，系统地介绍数据库的基本理论与实现技术。本书尤其适合与数据分析处理技术相关的非计算机专业人员使用，可帮助其跨越数据库技术门槛，掌握企业级数据分析处理方法，理解现代数据库技术的基本设计思想与技术发展趋势，了解商业智能的主要实现技术，熟悉现代企业数据分析处理的基本技术框架。本书也可以作为人文社会学科学生的数据库应用技术教材，建议在教学中侧重讲解基本数据库概念和 SQL 命令使用方法，掌握企业级数据库应用技术，可以相对弱化第 5 章数据库实现与查询优化技术与第 6 章数据库仓库和 OLAP 理论的要求，重点通过第 7 章 OLAP 实践案例掌握基于数据库的分析处理技术的完整数据处理流程，以及数据库分析处理的实用技能。

由于作者水平有限，编写时间仓库，书中难免存在疏漏和不足，恳请同行专家和读者给予批评和指正。

编著者

反侵权盗版声明

目　录

第 1 部分　导　论

第 2 部分　数据库基础知识与 SQL 实践

第 3 部分 数据仓库和 OLAP 基础

第1部分 导 论

本书以 Microsoft SQL Server 2017 为平台学习数据库分析处理技术，在导论中首先介绍 Microsoft SQL Server 2017 的安装与配置方法，以及通过 SQL Server 2017 导入和导出数据工具加载数据的方法。

对数据的分析处理是从数据中获取有价值信息的途径，数据分析主要采用表格、图表、地图等形式展现数据。当前比较有代表性的数据分析工具包括：Excel 数据透视图/表、Power Pivot for Microsoft Excel、Power BI Desktop、Tableau 等。这些数据分析工具易于使用，数据可视化功能强大，降低了用户数据分析的技术门槛。我们首先通过案例学习数据分析工具的基本用法，了解数据分析的终端需求，体会其背后的数据库基本理论的查询处理技术需求，为数据库理论的学习打下基础。

第 1 章　初识 SQL Server 2017

本章要点/学习目标

本章 1.1 节介绍 SQL Server 2017 在 Windows 平台的安装与配置；1.2 节介绍 SQL Server 2017 在 Linux 平台的安装与配置；1.3 节介绍 SQL Server 数据库数据导入和导出技术，通过案例实践让读者掌握数据导入数据库的不同方法；1.4 节介绍使用 Integration Services 导入数据的方法。

本章的学习目标是通过数据导入和导出案例介绍数据库加载功能，了解数据、结构、模式、数据类型的基本概念和特点，为学习数据库理论打下基础。

1.1　SQL Server 2017 在 Windows 平台的安装与配置

本节介绍 SQL Server 2017 在 Windows 平台的安装过程。SQL Server 是一个以数据库为中心的综合数据管理与分析处理平台，包括数据库引擎、Analysis Services、Integration Services、Report Services 等服务组件，支持包括数据库应用、OLAP 应用、数据挖掘应用和报表服务应用等不同层次的数据服务，与 BI 商业智能相结合，可以进一步支持可视化数据分析功能。

SQL Server 2017 提供了 Windows 平台和 Linux 平台版本，SQL Server 2017 需要独立安装 SQL Server 2017 数据库、SQL Server Management Studio 管理工具和 SQL Server Data Tools 数据集成工具。SQL Server 2017 可以从微软官方网站下载[1]，网站提供了 SQL Server 2017 on Windows、Linux、Docker 不同类型平台的下载版本供评估和开发使用。

下面以 Windows 10 平台上的数据库安装过程为例介绍 SQL Server 2017 数据库的安装步骤。

下载 SQL Server 2017 安装包后运行 SQL Server 2017 安装程序，在“SQL Server 安装中心”首先安装 SQL Server 2017 数据库引擎。在对话框左侧窗格中选择“安装”选项，执行“全新 SQL Server 独立安装或向现有安装添加功能”命令。

1）安装向导首先要求输入产品密钥。用户可以选择评估版本类型或选择其他版本并输入产品密钥，验证安装。

2）选择安装类型或输入正确的安装序列号后，确认接受许可条款。安装向导执行全局规则验证，确定在安装 SQL Server 程序支持文件时可能发生的问题，更正所有失败，保证安装程序继续进行。

3）安装程序执行 Microsoft 更新，也可以不勾选“使用 Microsoft Update 检查更新”复选框，跳过系统更新检查。然后安装向导开始扫描产品更新，下载安装程序文件，安装程序

---

1 https://www.microsoft.com/en-us/sql-server/sql-server-downloads

文件和安装程序文件过程如图 1-1 和图 1-2 所示。

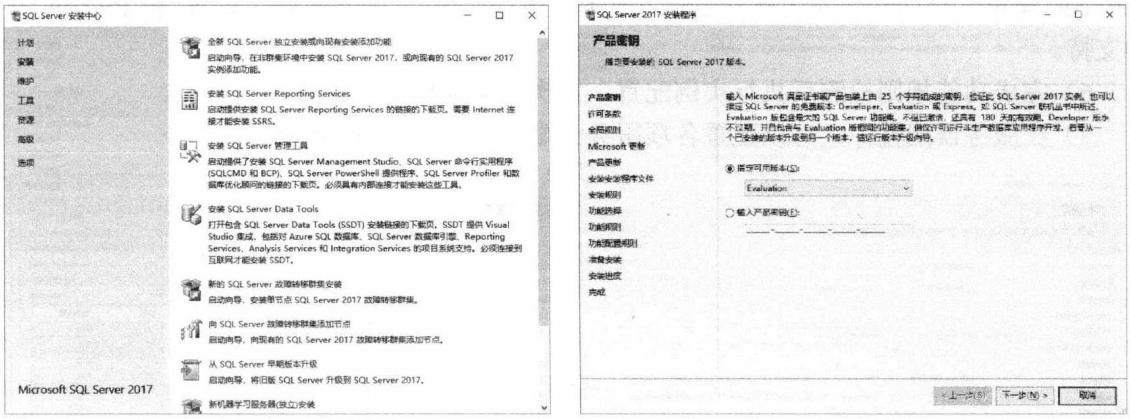

图 1-1

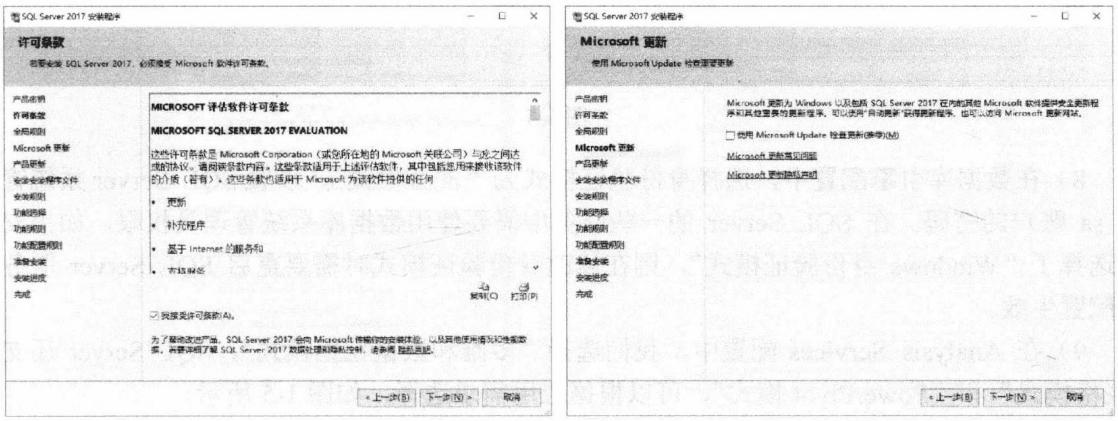

图 1-2

4）安装规则检测标识在运行安装程序时可能发生的问题，通过安装规则检测后才能继续后面的安装过程，如图 1-3 所示。

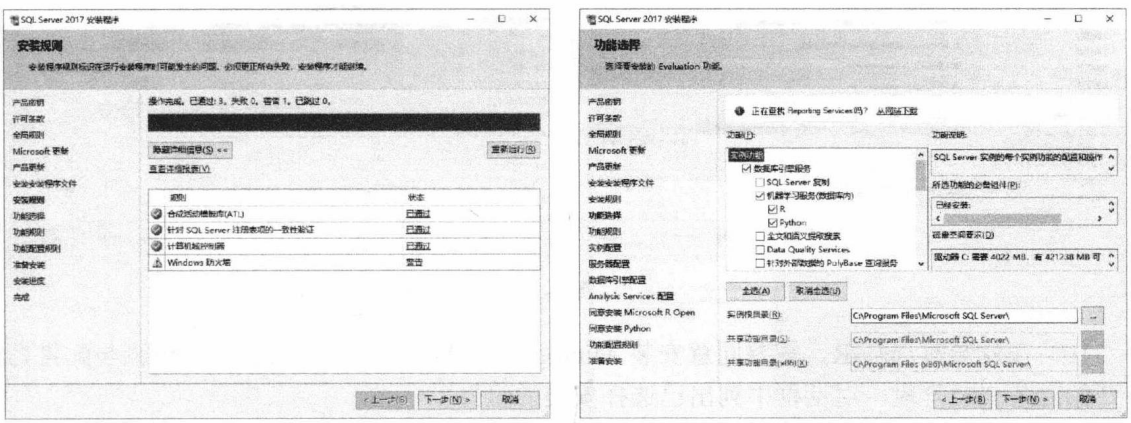

图 1-3

5）通过安装规则检测后，进入功能选择对话框，需要用户根据安装需求在功能窗口选

择 SQL Server 2017 相应的功能模块。在安装中需要选择“数据库引擎配置”“Analysis Services 配置”等功能，并选择机器学习服务（数据库内），安装数据库对 R 和 Python 语言的支持。

6）完成功能规则检测后执行实例配置，首次安装选择默认实例。

7）在服务器配置中，可以配置各项服务的账户信息，如图 1-4 所示。

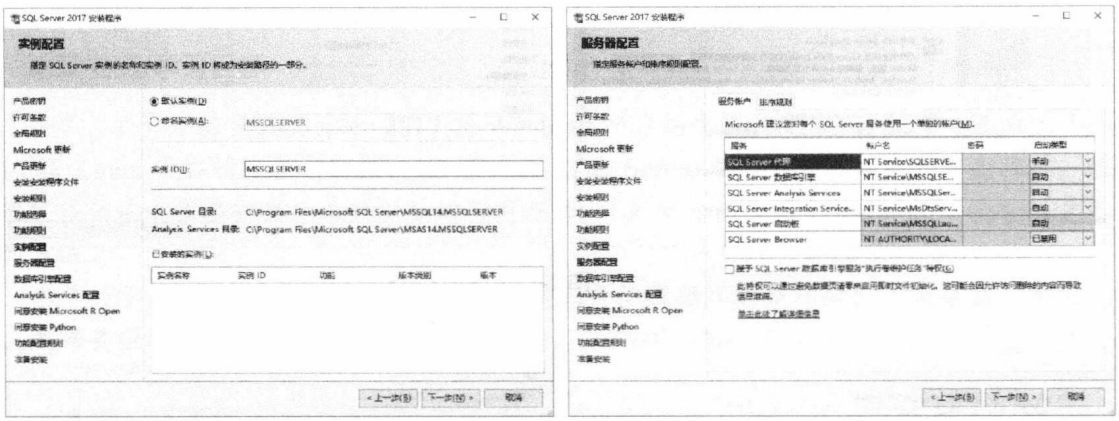

图 1-4

8）在数据库引擎配置中，选择身份验证模式为“混合模式”，设置 SQL Server 系统管理员 sa 账户的密码。在 SQL Server 的一些服务中需要使用数据库系统管理员权限，如果安装时选择了“Windows 身份验证模式”，则在修改身份验证模式时需要重启 SQL Server 服务以使配置生效。

9）在 Analysis Services 配置中，我们选择“多维和数据挖掘模式”，SQL Server 还支持“表格模式”和“PowerPivot 模式”，可以根据应用需求选择，如图 1-5 所示。

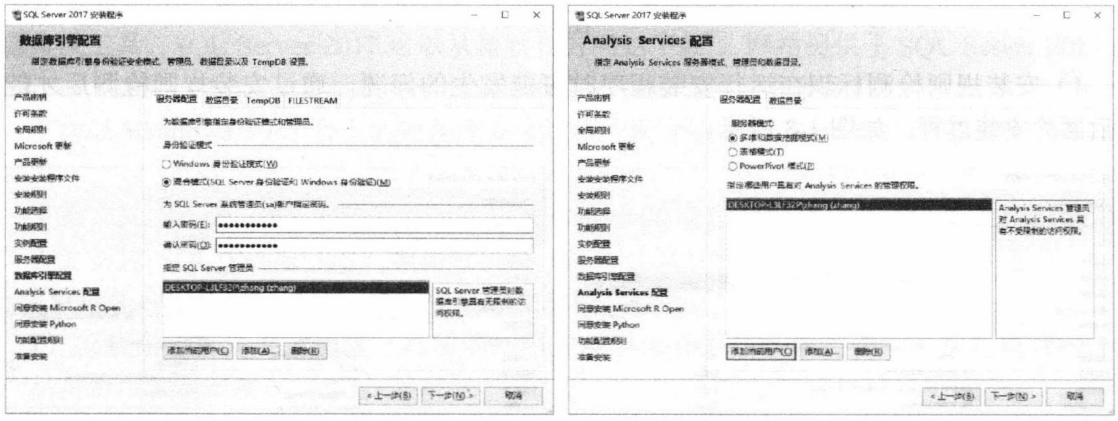

图 1-5

10）在安装时选择 R，需要同意安装 Microsoft R Open 协议。安装规则检测与配置完成后开始准备安装阶段，对话框中列出已选择安装的组件。

11）在安装 Python 时，需要同意安装 Python 及相关协议。

12）单击“安装”按钮后开始安装，安装程序对话框显示当前安装进度。当完成全部安装任务后，对话框显示完成状态，显示已成功安装的组件，如图 1-6 所示。

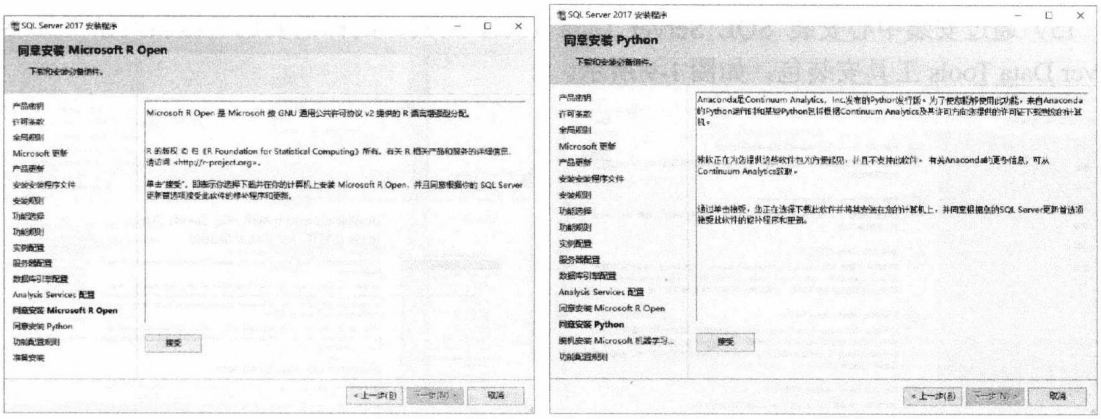

图 1-6

13）完成 SQL Server 引擎安装后再安装 SQL Server 管理工具，SQL Server Management Studio 需要从微软网站下载安装包，如图 1-7 所示。

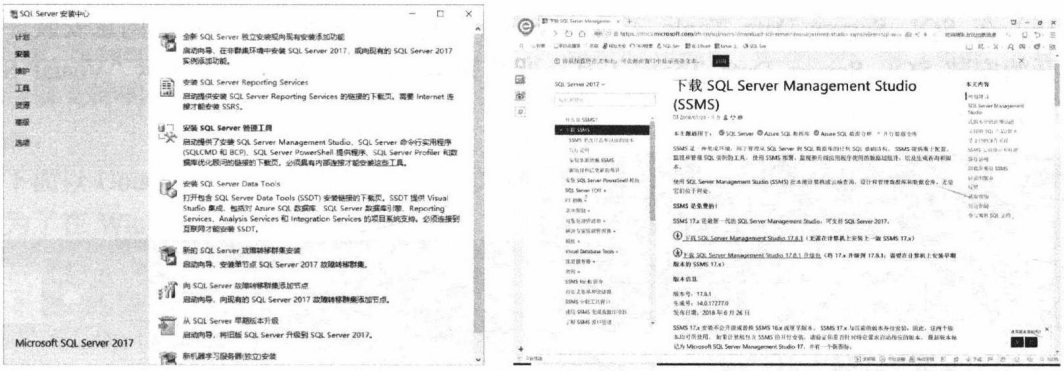

图 1-7

14）启动 SQL Server Management Studio 安装程序，根据安装向导完成安装。启动 SQL Server Management Studio 后显示 SQL Server 管理器，用于连接 SQL Server 引擎和使用查询器操作数据库中的数据，如图 1-8 所示。

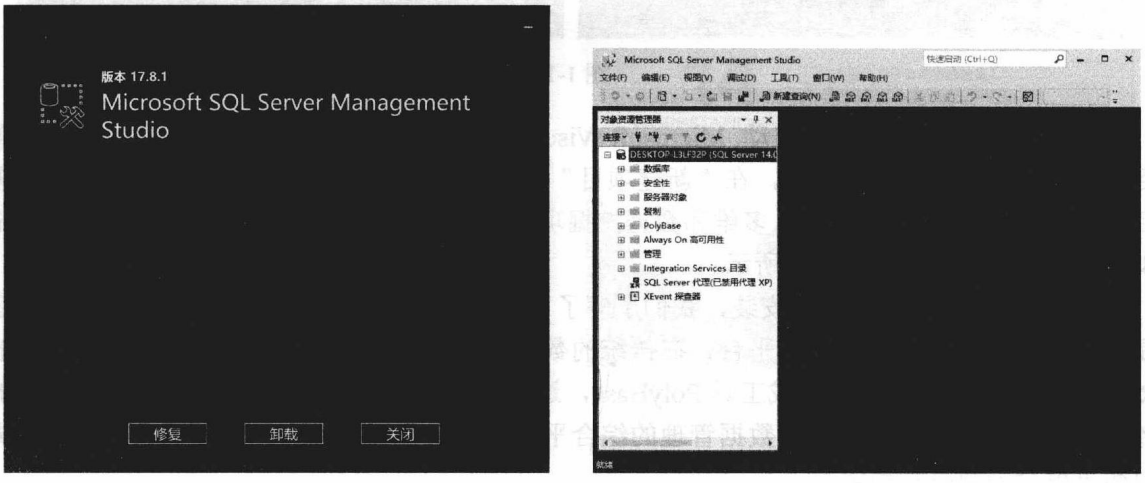

图 1-8

15）通过安装中心安装 SQL Server Data Tools 工具，同样需要通过微软网站下载 SQL Server Data Tools 工具安装包，如图 1-9 所示。

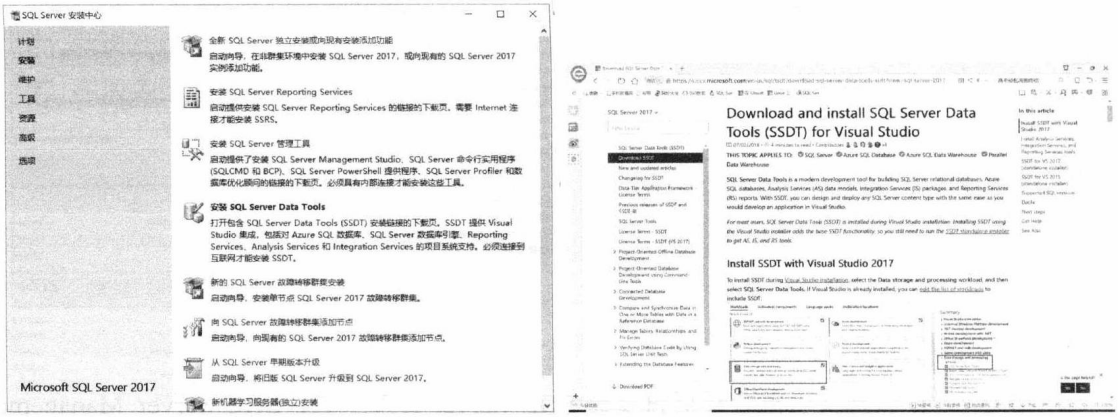

图 1-9

16）在 SQL Server Data Tools 工具安装向导中选择所需的工具，然后同意安装许可条款，开始安装 SQL Server Data Tools 工具。安装完毕后，在系统菜单中显示“Visual Studio 2017(SSDT)”，如图 1-10 所示。

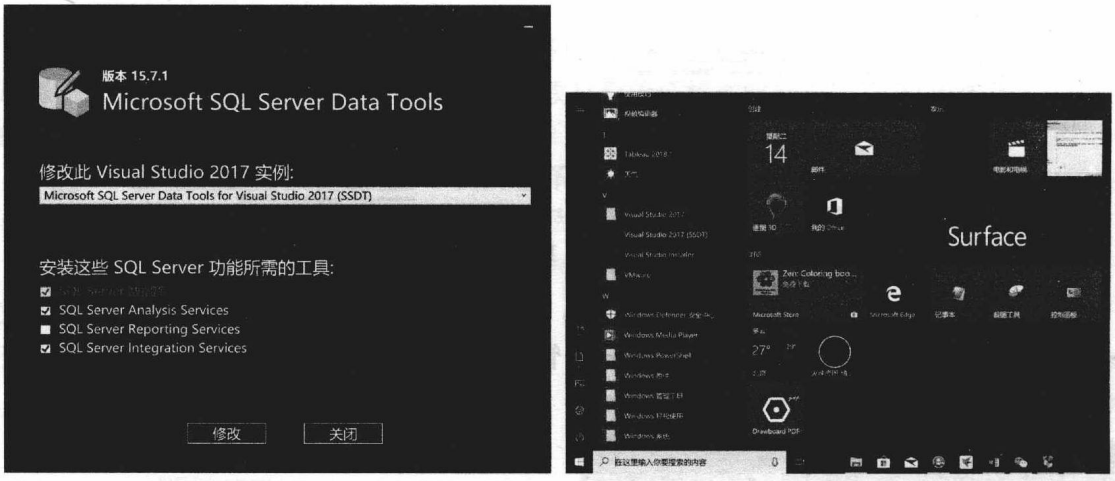

图 1-10

17）通过系统启动菜单启动 Microsoft Visual Studio，单击“文件”菜单中的“新建项目”下的“创建新项目”命令，在“新建项目”对话框中选择“商业智能”，可以看到新建项目对话框中 Analysis Services 多维和数据挖掘项目、Integration Services Project 和 Analysis Services 表格项目，如图 1-11 所示。

通过 SQL Server 2017 的安装，我们了解了 SQL Server 2017 不仅是一个数据库引擎，还是一个综合的数据管理与分析平台，在传统的数据库引擎基础上还集成了面向大数据分析的 R、Python 语言和 Hadoop 集成工具 PolyBase，这体现了当前和未来数据库产品和技术发展的趋势，即数据库逐渐成为一个数据管理的综合平台，面向不同结构的数据和数据管理平台提供数据融合与数据管理能力。