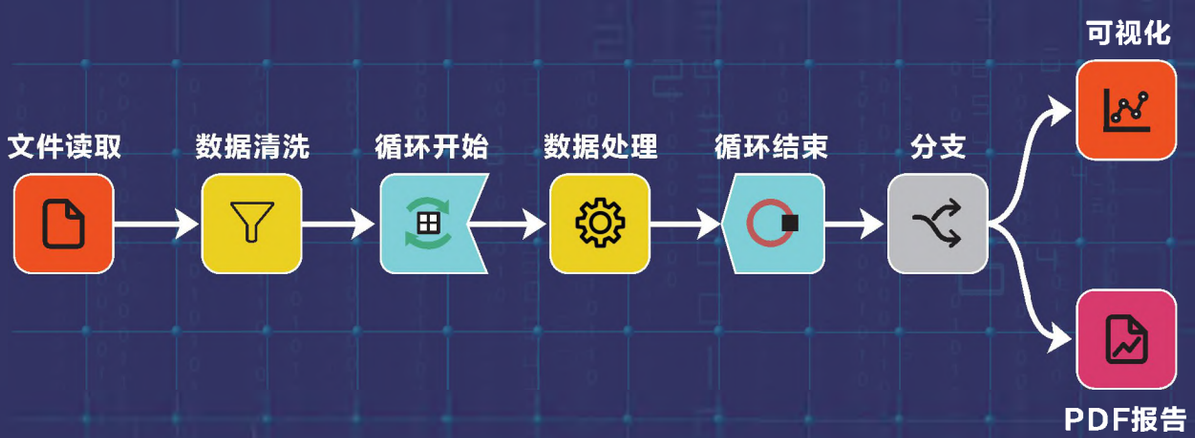
KNIME数据分析快速入门及进阶

赵永江／著

KNIME
数据分析
快速入门及进阶

赵永江／著





705分钟
同步视频讲解

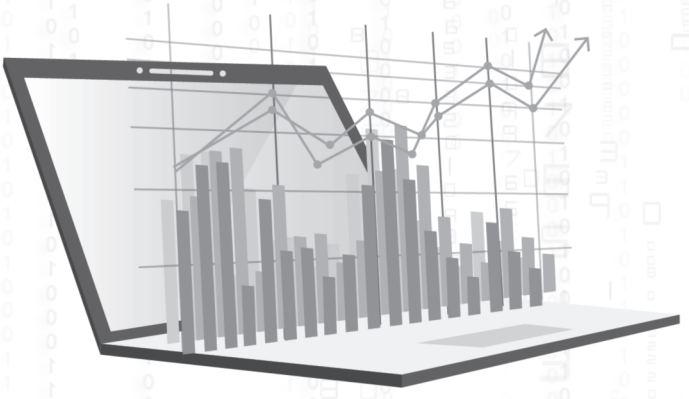
教学课件

配套工作流数据

电子科技大学出版社
University of Electronic Science and Technology of China Press

KNIME
数据分析
快速入门及进阶

赵永江／著



电子科技大学出版社
University of Electronic Science and Technology of China Press
·成都·

图书在版编目（CIP）数据

KNIME 数据分析快速入门及进阶 / 赵永江著. 成都 : 成都电子科大出版社, 2026. 2. -- ISBN 978-7-5770-2103-4

Ⅰ. TP274

中国国家版本馆 CIP 数据核字第 202577TV58 号

KNIME 数据分析快速入门及进阶

KNIME SHUJU FENXI KUAISU RUMEN JI JINJIE

赵永江　著

策划编辑　刘亚莉
责任编辑　刘亚莉
责任校对　岳寒晨
责任印制　梁　硕

出版发行　电子科技大学出版社
　　成都市一环路东一段 159 号电子信息产业大厦九楼　邮编 610051
主　　页　www. uestcp. com. cn
服务电话　028-83203399
邮购电话　028-83201495

印　　刷　成都市东辰印艺科技有限公司
成品尺寸　210 mm×297 mm
印　　张　23. 5
字　　数　700 千字
版　　次　2026 年 2 月第 1 版
印　　次　2026 年 2 月第 1 次印刷
书　　号　ISBN 978-7-5770-2103-4
定　　价　98. 00 元

版权所有，侵权必究

前　言

近年来，数据分析领域经历了前所未有的变革和发展。数据的生成速度和规模达到了史无前例的高度，企业也积累了越来越多的数据，迫切需要员工分析大量数据。然而，当前面临这样一个困境：一方面，传统的电子表格软件在处理海量数据时显得力不从心，常常崩溃且难以实现复杂的自动化流程；另一方面，Python、R 等编程语言学习成本高，又将一些人挡在了数据科学的大门之外。是否存在一座既连接业务知识又跨越技术鸿沟的桥梁呢？答案是肯定的，这座桥梁就是 KNIME。

KNIME Analytics Platform 是一款开源、免费的数据科学工具，以直观、可视化的方式改变了我们与数据交互的模式。KNIME 无须编写代码，只需拖曳并连接不同任务的节点即可将其组合成一个逻辑清晰的工作流。无论是数据的读取、清洗、转换，还是构建复杂的机器学习模型，整个分析过程都如同流程图般一目了然，可谓是思维导图式分析。这种可视化操作，极大地降低了数据分析的门槛，使财务、市场、运营、研发等各个方面的专家拥有了独立解决数据问题的能力。他们可以借助 KNIME 在几分钟之内完成过去需耗时数小时甚至数天的数据处理任务，从而将宝贵的精力投入到更具价值的洞察与决策中。

KNIME 直观易学、功能强大、灵活性高，赢得了全球众多行业领导者的信赖。公开资料显示，宝洁公司将其应用于供应链管理；西门子公司利用 KNIME 自动处理一些烦琐的任务，以提高工作效率；花旗银行将其作为企业级分析平台，稳定运行着上千个关键业务工作流；希捷科技公司将其应用于业务、研发与制造环节中。在国内，KNIME 也被应用于各行各业，如金融、审计、快消品、医药等。同时，国内一些高校也逐渐开设了关于 KNIME 的课程。

此外，KNIME 还是一个可商用的开放平台，持续集成最新技术，从大数据到人工智能，始终走在技术前沿。KNIME 能够与 Python、R、Java 等编程语言和 Hive、Spark 等大数据平台无缝集成，扮演着“编排者”的角色，让不同技术背景的团队成员在 KNIME 平台中进行高效协作。

自 2020 年起，为了让更多人了解和掌握 KNIME，我曾以“不编程亦分析”为网名（现更名为“玩转 KNIME”），在哔哩哔哩等平台上分享 KNIME 的学习方法与实战技巧。之所以使用这个网名，意在强调即使不会编程，也能做好数据分析。在分享知识过程中，我收到了众多 KNIME 爱好者的鼓励和支持，也收到许多求助和交流的私信。倍感荣幸的是，我曾受邀到某世界五百强企业总部进行现场培训。我也曾在多家世界五百强企业总部工作过，十多年来一直使用 KNIME 解决实际业务中的各种问题，同时在公司内部分享如何学习和应用 KNIME。此外，我也获得了 KNIME 官方认证培训师资格（KNIME Certified Trainers）。这些经历让我萌生了一个想法：或许应该撰写一本既通俗易懂又能贴近实战的有关 KNIME 的书，让更多人体会数据分析的乐趣。

很多人可能会问，学 KNIME 需要什么基础吗？我非常肯定地告诉大家：会用 Excel 基本函数，学习 KNIME 就完全没问题。只要你下定决心学习 KNIME，本书就可以带你快速入门，掌握 KNIME 的核心技能。也有人会问，KNIME 适用哪些行业？实际上，KNIME 是一个通用的数据分析工具，并不局限于特定行业，具体用途取决于用户的需求和使用场景。

本书是一本以 KNIME 为工具，以实战为导向的技术类图书。本书遵循由浅入深的原则，精心编排各章知识点和相关示例，推荐初学者按章节顺序学习。下面介绍本书的主要内容，全书分为 6 大部分，共计 18 章。

第 1 部分为 KNIME 软件的安装及必备基础，包括第 1 章和第 2 章，主要涵盖 KNIME 简介及安装、扩展的安装及删除、KNIME 必备基础和操作、核心概念等。

第 2 部分为数据分析必知必会内容，包括第 3 章至第 8 章，主要涵盖数据分析核心技能。①数据读写：掌握文件合并、路径设置及读写节点的操作；②基础操作：包括行列过滤、排序、类型转换；③条件判断：包括多条件应用及节点设置；④表级别操作：重点学习合并、切分、结构重塑（长宽表转换、行列互换等）及表格属性设置和管理；⑤数据汇总：以 GroupBy、Pivot 节点为核心，结合案例掌握汇总三要素及移动窗口分析；⑥可视化：学习图形节点分类、应用场景，为生成报告和掌握仪表盘的使用方法奠定基础。

第 3 部分为数据分析进阶必备内容，包括第 9 章至第 12 章，主要涵盖高级数据处理与自动化技术。①流变量：学习变量类型、创建与使用，了解流变量是循环、分支等模块的基础；②循环：解析循环四要素，掌握不同类型的循环、循环嵌套等复杂逻辑；③分支与错误处理：通过分支四要素框架和自动发送邮件的案例，实现条件判断与异常管理；④字符串高阶操作：重点学习正则表达式应用（匹配、脱敏、加密等）、字符串相似度比较及数据标准化。

第 4 部分为数据分析进阶应用，包括第 13 章至第 16 章，各章节为 Component 组件的不同场景应用。①报告：解析报告七元素与布局调整，实现一份或多份报告生成；②仪表盘：学习交互设计、参数关系，了解 Configuration 与 Widget 节点的区别；③自定义组件：包括简化和封装工作流、总结组件五大应用场景；④流式执行：学习通过封装组件来优化执行逻辑，加快数据处理速度。

第 5 部分为利用编程工具增强 KNIME 的功能，位于本书第 17 章。本部分主要介绍 Python 集成环境和用户自定义虚拟环境，如何在 KNIME 中应用虚拟环境编写 Python 代码等。

第 6 部分为精选案例及思考总结，位于本书第 18 章。本部分通过精选来自投资管理、稽核审计、业务管理等领域的案例，展示数据处理和分析过程，强化对循环、组件、流变量等知识点的综合运用，并对 KNIME 的学习进行总结。

为便于读者学习，本书配套各章同步学习视频、KNIME 工作流（包含数据）、教学课件、教学大纲等内容，请扫描下方的二维码获取。

在写作本书的过程中，我得到家人的大力支持和协助，感谢他们的默默付出和支持。此外，感谢第一个向我推荐 KNIME 的比利时同事蔡骏博士，以及在工作中给予我指导的上海财经大学罗明博士；也特别感谢 KNIME 官方教育主管 Schalk Gerber 博士在 KNIME 教学方面给我的指导，以及北京市盈科（无锡）律师事务所党支部书记薛孝东律师和无锡市大数据协会的鼎力支持。最后，衷心感谢电子科技大学出版社编辑给予的专业指导与宝贵建议。同时，也向出版社所有参与本书编校、排版及设计工作的老师表示深深的谢意。

本书所涉及的人名、公司名均为作者在写作过程中所虚构的，仅用于说明、演示或分析相关内容，不构成任何现实主体的价值判断、不当关联或法律评价。

限于个人水平，书中难免存在疏漏之处，欢迎广大读者批评指正。错误反馈或建议请发邮件至 yongjiang2312225@ 163. com。

作　者

2025 年 10 月

目　录

第 1 章　KNIME 简介及安装

1.1　KNIME 主要功能 ……（ 2 ）
1.2　KNIME 产品体系 ……（ 3 ）
1.3　KNIME 安装 ……（ 3 ）
1.4　KNIME 软件设置 ……（ 6 ）
1.5　KNIME 扩展安装及卸载 ……（ 7 ）
1.6　KNIME 两种界面的比较 ……（ 11 ）

第 2 章　KNIME 必备基础知识

2.1　节点（Node） ……（ 18 ）
2.2　工作流（Workflow）及工作流组（Workflow Group） ……（ 23 ）
2.3　工作流注释 ……（ 28 ）
2.4　数据类型及其转换 ……（ 29 ）
2.5　核心概念 ……（ 30 ）

第 3 章　数据分析概览及数据读写

3.1　数据分析流程简介 ……（ 32 ）
3.2　数据结构及数据来源 ……（ 34 ）
3.3　常用数据读入类节点 ……（ 35 ）
3.4　常用数据写入类节点 ……（ 46 ）
3.5　文件路径 ……（ 49 ）

第 4 章　基础数据操作

4.1　过滤 ……（ 54 ）
4.2　排序 ……（ 59 ）
4.3　列重命名 ……（ 62 ）
4.4　数据类型转换 ……（ 64 ）
4.5　常见数据类型的操作 ……（ 66 ）

第 5 章　条件判断

5.1　单条件判断 ……（87）
5.2　多条件判断 ……（90）
5.3　复杂条件判断编程类节点 ……（97）

第 6 章　表级别操作与属性管理

6.1　表格左右合并 ……（102）
6.2　表格左右切分 ……（108）
6.3　表格上下合并 ……（109）
6.4　表格上下切分 ……（110）
6.5　表格裁切 ……（113）
6.6　表格结构重塑 ……（114）
6.7　获取表格属性 ……（122）
6.8　修改表格属性 ……（123）
6.9　验证表格属性 ……（128）
6.10　比较表格属性或数据差异 ……（130）
6.11　重要节点小结 ……（131）

第 7 章　数据汇总

7.1　汇总三要素 ……（134）
7.2　汇总类核心节点——GroupBy ……（137）
7.3　透视汇总节点——Pivot ……（142）
7.4　行列汇总 ……（144）
7.5　移动窗口汇总 ……（147）
7.6　集合类型数据聚合 ……（150）

第 8 章　数据可视化

8.1　可视化节点概述 ……（155）
8.2　数据准备 ……（158）
8.3　基于维度的可视化节点分类 ……（160）
8.4　可视化节点的设置 ……（167）
8.5　绘图结果的应用 ……（171）

第 9 章　流变量

9.1　流变量的类型及端口 ……（175）
9.2　创建和查看流变量 ……（179）
9.3　编辑和使用流变量 ……（185）

9.4 流变量与组件（Component） ……（193）

第 10 章 循环

10.1 循环概述 ……（198）
10.2 计数循环 ……（200）
10.3 基于行分组的循环 ……（202）
10.4 基于列分组的循环（Column List Loop Start） ……（205）
10.5 条件循环（Generic Loop Start） ……（208）
10.6 递归循环（Recursive Loop Start） ……（211）
10.7 多个循环间的关系 ……（218）
10.8 循环流程的控制 ……（222）
10.9 常见循环问题 ……（224）
10.10 带有循环功能的其他节点 ……（227）

第 11 章 分支与错误处理

11.1 分支四要素 ……（230）
11.2 两分支 ……（232）
11.3 多分支 ……（234）
11.4 分支与循环 ……（242）
11.5 其他扩展分支 ……（243）
11.6 错误处理（Try 和 Catch Errors） ……（247）

第 12 章 字符串高级操作

12.1 正则表达式基础 ……（253）
12.2 KNIME 正则节点及应用 ……（260）
12.3 易错易混正则知识点 ……（266）
12.4 其他高阶操作 ……（268）

第 13 章 报告

13.1 创建报告流程 ……（276）
13.2 报告七要素 ……（279）
13.3 批量创建报告 ……（289）
13.4 报告中使用非兼容可视化节点 ……（291）

第 14 章 动态数据应用：交互式仪表盘

14.1 创建交互式仪表盘步骤 ……（294）
14.2 Component 参数及视图布局 ……（295）
14.3 特殊交互式控件 ……（304）

14.4 Widget 类和 Configuration 类节点比较 ……（306）
14.5 报告与动态仪表盘的区别 ……（310）

第 15 章 自定义图形化用户界面

15.1 创建自定义节点 ……（312）
15.2 管理自定义节点 ……（314）
15.3 创建图形化用户界面工作流 ……（319）
15.4 Component 总结用途 ……（320）

第 16 章 组件的流式执行

16.1 流式执行概述 ……（322）
16.2 处理超大单个文件 ……（324）
16.3 流式执行和顺序执行的区别 ……（326）

第 17 章 利用 Python 赋能 KNIME

17.1 Python 集成环境 ……（328）
17.2 Conda 创建 Python 虚拟环境及其应用 ……（331）
17.3 Pixi 创建 Python 虚拟环境及其应用 ……（339）

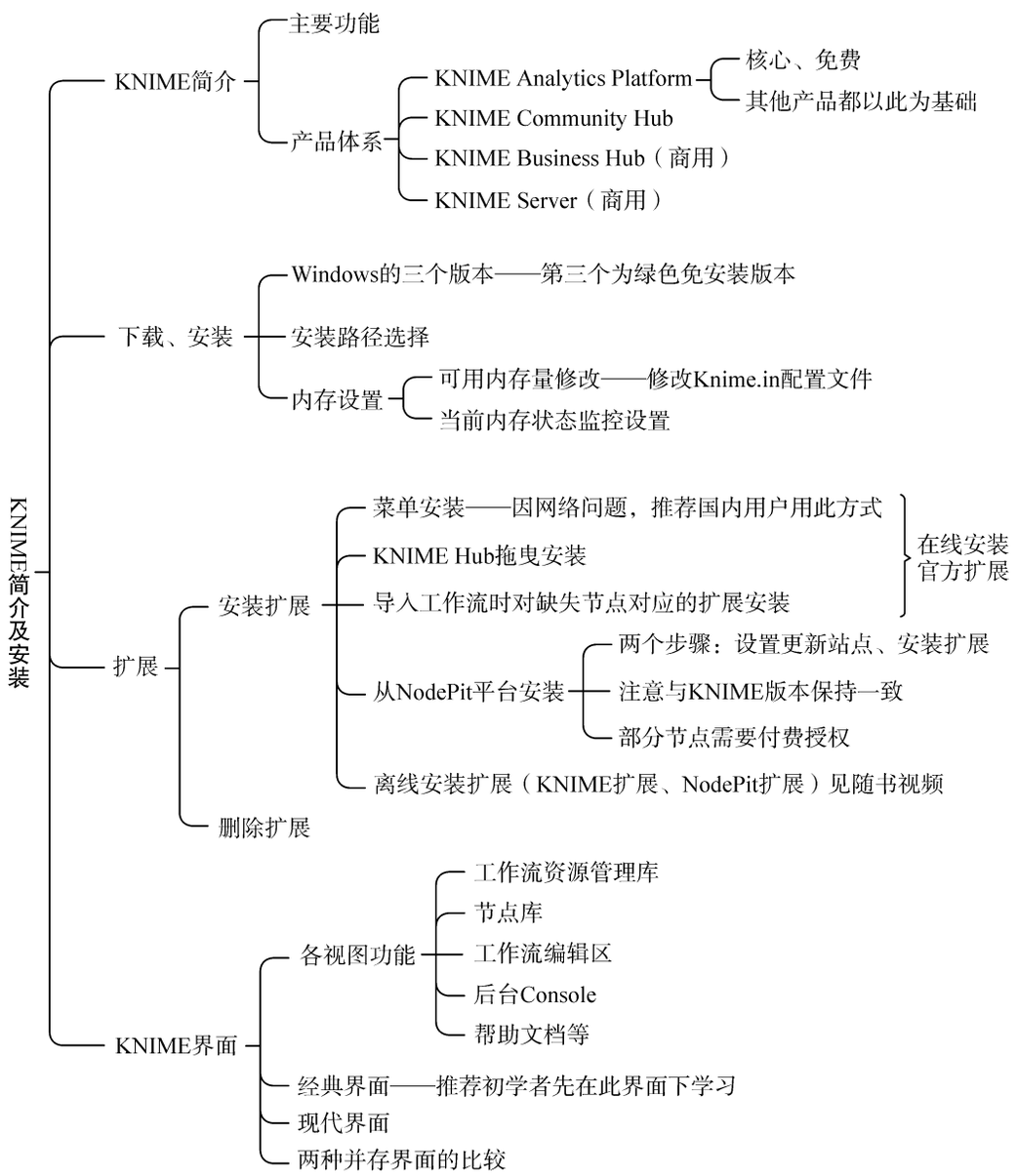
第 18 章 实战案例与技巧分享

18.1 计算连续工作日 ……（343）
18.2 机房基站全部故障识别及持续时间统计 ……（346）
18.3 串标、围标风险识别 ……（350）
18.4 债券信用评级有效起止时间段计算 ……（352）
18.5 领料单数据汇总 ……（355）
18.6 Excel 数值转日期 ……（359）
18.7 KNIME 学习的思考与总结 ……（362）

附录：常用节点名称变化对照表 ……（365）

第 1 章　KNIME 简介及安装

本章将介绍 KNIME 的主要功能，软件安装和设置，扩展的安装和删除，介绍 KNIME 5.0 及之后版本中差异较大的两类并存界面（经典界面和现代界面），总结两者差异和优缺点，并为初学者提供学习建议。首次学习 KNIME 的读者在阅读本书的过程中可能对有关概念或关键词不理解，可快速阅读本章，在学习第 2 章基础知识后再回顾本章内容。本章主要知识点总结如下。



1.1 KNIME 主要功能

KNIME（Konstanz Information Miner，康斯坦茨信息挖掘器）Analytic Platform 简称 KNIME 或 KNIME AP，是一款免费、开源的数据科学分析平台软件，该软件最早是德国康斯坦茨大学（Universität Konstanz）的研究项目，现已发展为一款流行的商业化数据分析平台。由于其直观、易用、功能丰富等特点，被业界誉为数据科学界的“瑞士军刀”。KNIME 通过拖曳式操作，无须编程即可快速实现数据处理和分析，简单直观易用，适合各行各业的数据分析人员，尤其是重度使用 Excel 且不会编程的用户。

在 KNIME 中进行数据处理与分析、机器学习等，与画流程图类似，本书形象地称其为思维导图式数据分析。图 1-1 中的每个方块代表一个节点，可视为一个数据处理环节，每个步骤数据都可随时查看，过程清晰易懂，实现了一次创建、多次使用。该流程图还实现了在一个表内多个 Sheet 表（工作表）合并的过程，仅需设置第一个和第三个节点，无须任何代码，拖曳式操作，一分钟内即可快速完成数据合并。

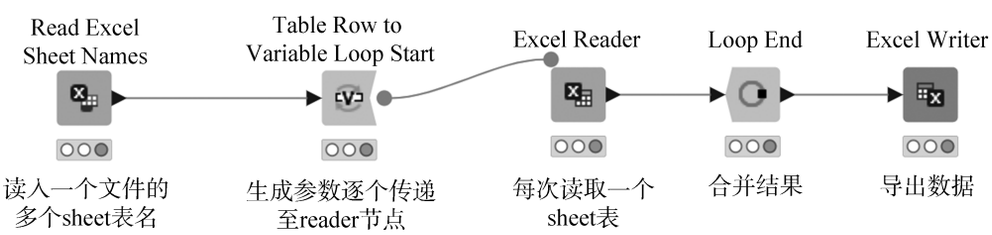


图 1-1　利用 KNIME 实现一个工作簿多个 Sheet 表的合并

KNIME 是一个强大的数据分析工具，它可以从各种数据源（如数据文件、数据库等）中读取数据。将数据导入后，KNIME 提供了丰富的功能来清洗和转换数据、数据可视化、建立和评估各种机器学习模型、输出格式化报告、发送邮件等，以实现数据处理和分析全流程工作自动化。直观易懂的可视化操作，极大简化了日常性的烦琐工作，提高了工作效率。KNIME 支持与其他编程语言协作，如现有节点和扩展不能满足自己的特殊需要，可利用主流的编程语言实现某一特殊功能。KNIME 支持 Java、JavaScript、R、Python、Ruby 等编程语言，可根据需要调用第三方软件包，实现所需特殊功能。KNIME 支持各种主流机器学习建模（如 Scikit-learn、TensorFlow）和大数据平台的工具（如 Spark、Hive 等）。丰富的扩展包和广泛的适用性，使 KNIME 能够为用户提供一站式全流程数据处理和分析。

KNIME 功能丰富，按照其主要功能大致概括为以下六个方面。

（1）数据读入和处理。KNIME 支持多种不同格式数据文件，如常见的 Excel、CSV 文件；也支持直接连接的各种数据库，如 MySQL、Hive、MongoDB。在 KNIME 中读入数据之后，可以进行数据清洗、转换、汇总等，实现数据的自动化、流程化。对于烦琐的日常工作，一次创建，多次使用，该过程常被视为 ETL（Extract 抽取、Transform 转换、Load 载入）过程。

（2）模型的构建、评估、优化。KNIME 支持各种机器学习算法，如常见的分类、回归、聚类、关联规则、神经网络等，并提供模型评估和参数调优的简单、易用节点，使得用户无须拥有编程基础也可以构建复杂的机器学习模型。

（3）数据可视化。KNIME 提供了各种开箱即用的可视化图形，以直观的方式展示数据，实现对数据的更深理解，为后续的探索分析提供更多的支持。此外，也可实现各种高度自定义图形，如基于 JavaScript 的开源可视化图表库 Apache ECharts，以及 Python 图形库等。

（4）工作流自动化。根据 KNIME 提供的各种功能节点，结合业务逻辑和规则，将数据分析流程固化于一个个有机地连接在一起的数据处理流程中，形成一个完整的分析流程，每一个步骤结果都可随时被查看。尤其对于重复性的工作，如对日报、周报、月报等的数据分析，可一次创建，多次重复使

用。同时，KNIME 还支持自动化执行和批处理，定时启动任务，提高工作效率。

（5）支持多种编程语言和丰富的扩展。KNIME 基于 Java 编写，因此支持 Java 各种函数和模块，又支持安装其他编程语言扩展，如比较流行的 Python、R、JavaScript 等编程语言。可利用编程语言简化部分复杂步骤，实现不同工具的融合与协作。KNIME 还提供诸如分子生物学、医药学等专业领域的扩展。

（6）支持团队协作和快速构建业务应用。基于 KNIME 构建的工作流发布在 KNIME Business Hub 中，可与团队共享和协作，还可在 KNIME 服务器端快速部署业务应用和服务，而无须拥有前后端的专业知识。（注意，KNIME Business Hub 为基于 KNIME 实现协作部署的商业软件，本书不涉及。）

1.2　KNIME 产品体系

随着 KNIME 的发展，其配套产品名称和功能有所变化，本节将简要介绍各产品的通用名称和功能，但一般用户常用前两个产品，最后两个为商业付费产品，简单了解即可。

（1）KNIME Analytics Platform，一般简称为 KNIME AP 或 KNIME，是免费、开源的数据分析平台软件。该软件是其他产品或服务的基础。在 KNIME 软件中，可通过网络将构建的工作流发布到 KNIME Community Hub 和 KNIME Business Hub 中。

（2）KNIME Community Hub，一般简称为 KNIME Hub，它提供了各种共享的节点、组件、工作流，也可进行小组内的共享协作。简言之，其功能类似编程人员常使用的 GitHub。除了在 KNIME 软件中能找到 KNIME Hub 标签页外，也可通过浏览器访问 KNIME Hub。在 KNIME Hub 的基础上，KNIME 又推出了 KNIME Pro、KNIME Team 等衍生产品，这两者都是基于云端的应用，可实现在线自动化、部署工作流、协作等。KNIME Pro 定位个人专业用户，KNIME Team 定位小型团队，普通读者简单了解即可。

（3）KNIME Business Hub，是基于可扩展的云架构的商业产品，需要额外安装，用于工作流的协作和共享。它是将 KNIME 构建的本地工作流以数据应用和服务的形式，快速部署于远程服务器端，并对工作流进行监控和版本管理。相对于 KNIME Server 来说，它在协作、易用性、部署、监控、许可管理等方面做了重大改进。

（4）KNIME Server，是早期开发的服务器端产品，可提供远程工作流的执行、协作和监控，以及用户管理操作；但具有用户权限管理不便和用户数量存在上限等缺点。根据官方的规划和确认，该产品最终可能由 KNIME Business Hub 替代。

KNIME 中前三个产品的关系如图 1-2 所示，本书聚焦于讲解 KNIME AP。

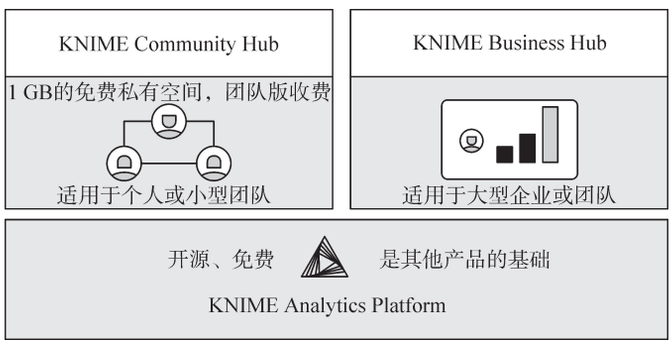


图 1-2　KNIME 三个产品的关系

1.3　KNIME 安装

如图 1-3 所示，打开 KNIME 官网，在网站右上角单击 Download，页面将跳转至产品注册信息页

面，在该页面内可不填写信息，勾选最底部复选框，再单击页面下方的 Download。此时将跳转至下载页面，如图 1-4 所示。该页面提供了 Windows、Linux、Mac 三种不同版本的安装包，根据系统选择相应的安装包。

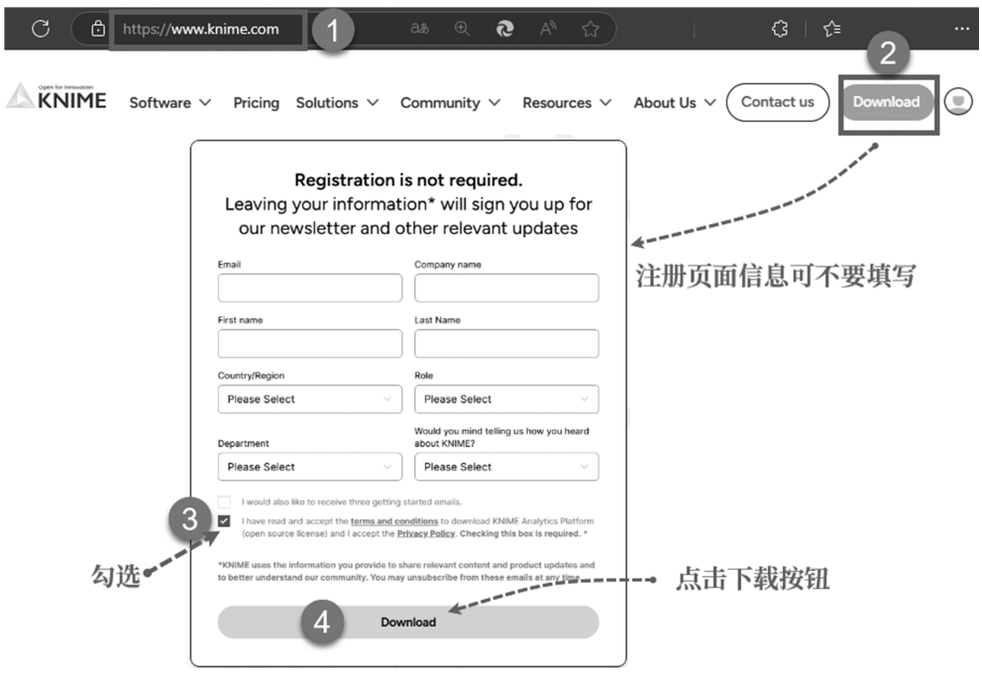


图 1-3　KNIME 官网下载地址

自 KNIME 5.5 版本起，KNIME 的发布策略有了新的调整，引入了两种并行的发布版本，标准版（Standard Release）和长期支持版（Long-Term Support）。标准版自 5.5 版本开始每 6 周更新 1 次版本，目前是 5.7 版本。长期支持版目前为 5.5.2 版本，大约 1 年提供 2 次更新。建议初学者先下载长期支持版安装包。Windows 格式有三种不同的安装包，如图 1-4 所示，依次为常用安装包、自解压安装包（无须第三方解压软件）、绿色免安装包。初学者建议选择第一种安装包，因为在安装过程中，它会根据电脑的硬件内存大小进行内存大小动态配置。对于第三种安装包，解压后在文件夹中双击 knime.exe 即可执行。第三种安装包有两个好处：第一，若本机已使用第一种安装包安装了 KNIME 5.4，可下载标准版的 Zip 压缩文件，解压文件后双击文件夹内的 knime.exe，借此可以实现一机多版本共存；第二，若在本机中用户没有安装软件的权限，直接使用绿色免安装版，跳过软件安装步骤。



图 1-4　KNIME 软件下载页面

下载 Windows 安装包执行文件后，双击安装。安装过程如图 1-5 和图 1-6 所示。



图 1-5　KNIME 安装过程（一）

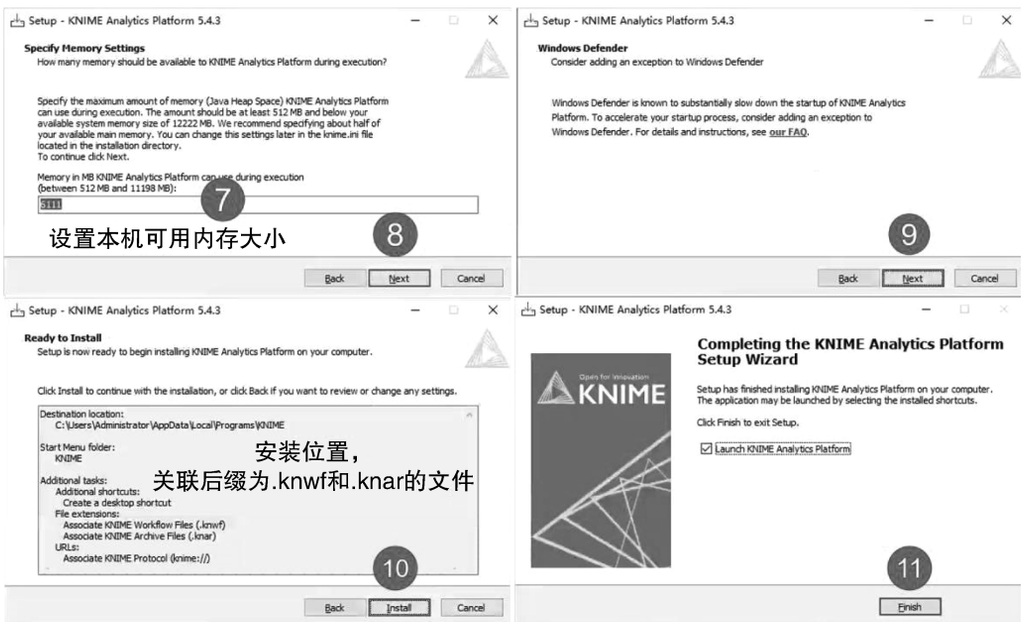


图 1-6　KNIME 安装过程（二）

首次安装 KNIME 时，默认安装在 C 盘位置，可更换指定目录至其他位置。在图 1-5 标号 6 处，默认勾选全部复选框，关联后缀为. knwf 和. knar 的文件。这两种类型的文件都为 KNIME 所创建的文件。最后，关联 KNIME 协议，即支持 KNIME 协议类型文件路径，该路径在本质上是相对路径的一种表示形式，本书后续章节会提及。

在安装过程中，KNIME 将根据本机内存的大小，分配约 50% 的可用内存量。若本机内存硬件升级，则可在安装目录位置中找到 knime. ini 配置文件，如图 1-7 所示；在该配置文件中将-Xmx16246 m 的数值修改为更大的数值，以提高 KNIME 软件可用内存量。若使用绿色版 KNIME，请手动修改配置文件中可用内存大小。

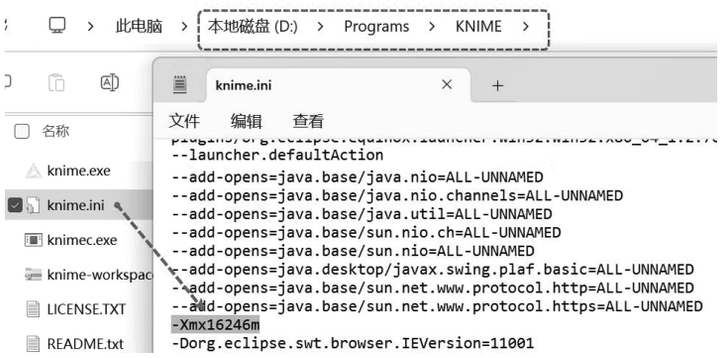


图 1-7　修改 KNIME 配置文件中可用内存大小

1.4　KNIME 软件设置

1.4.1　界面切换设置

首次启动 KNIME 后将弹出窗口，询问用户 KNIME 后续创建工作流文件要保存在哪个文件夹，如图 1-8 所示。若不想每次启动软件都跳出该窗口，可勾选底部不要再询问复选框选项，后续所有创建工作流文件都将保存在此路径下。建议将该目录设置在非系统盘位置，这样即使重装了 Windows 系统，此前 KNIME 创建的文件还在此文件夹中，不会因此丢失数据。

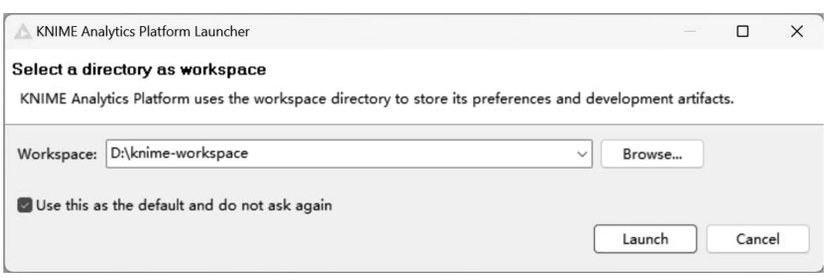


图 1-8　KNIME 首次启动时文件保存路径设置窗口

首次启动 KNIME 的界面如图 1-9 所示，请单击右上方菜单（Menu→Switch to classic user interface），将 KNIME 的界面从默认的现代界面 Modern UI 切换到经典界面 Classic UI（User Interface 用户界面）。之所以先切换至经典界面，是因为从 5.0 版本开始，KNIME 软件界面做了重大界面创新，并存在两种不同的用户界面，而经典界面更适合初学者。

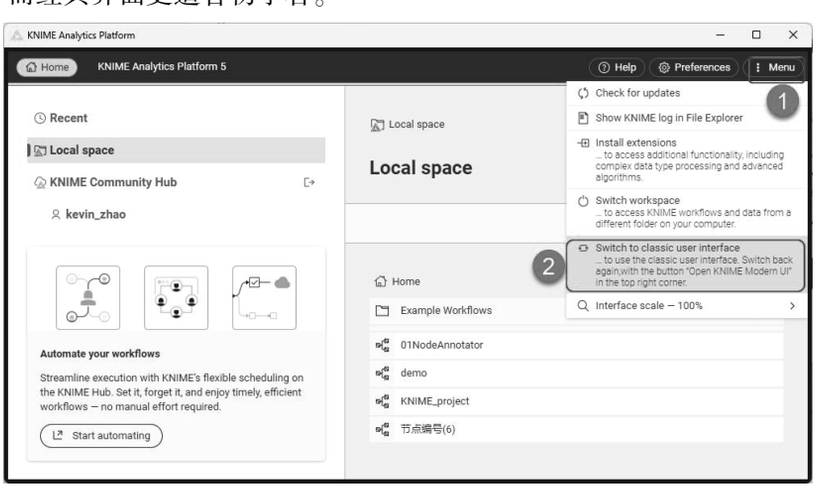


图 1-9　首次启动 KNIME 界面（默认现代界面）及切换设置

Classic UI 为经典界面，如图 1-10 所示。Modern UI 为革命性的现代界面。由于后者做了很多减法，界面过于简洁，甚至部分功能缺失，建议初学者先切换至经典界面学习。如果是经典界面的老用户，切换至现代界面也能很快适应。因此，本书如无特别说明，软件操作一般采用经典界面。

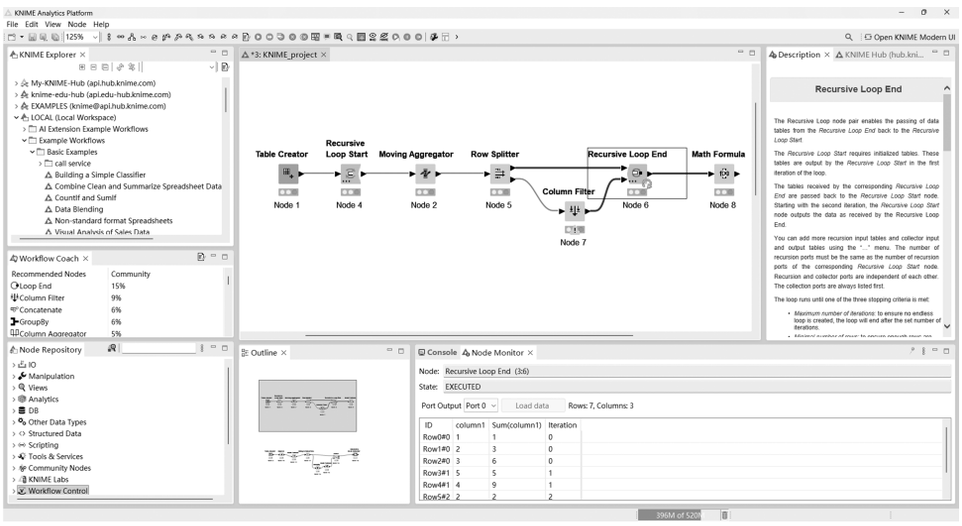


图 1-10　经典界面 Classic UI

1.4.2　当前内存使用显示设置

KNIME 切换至经典界面后，点击 File→Preferences→General，并勾选复选框 Show heap status 选项，如图 1-11 所示。设置完成后，在软件界面的右下方将看到当前 KNIME 软件内存的使用情况。该设置的主要目的是，实时监控 KNIME 当前内存使用量。当运行大量数据时，内存使用量接近安装时设置的内存上限，将会导致运行变慢，甚至软件崩溃或闪退。

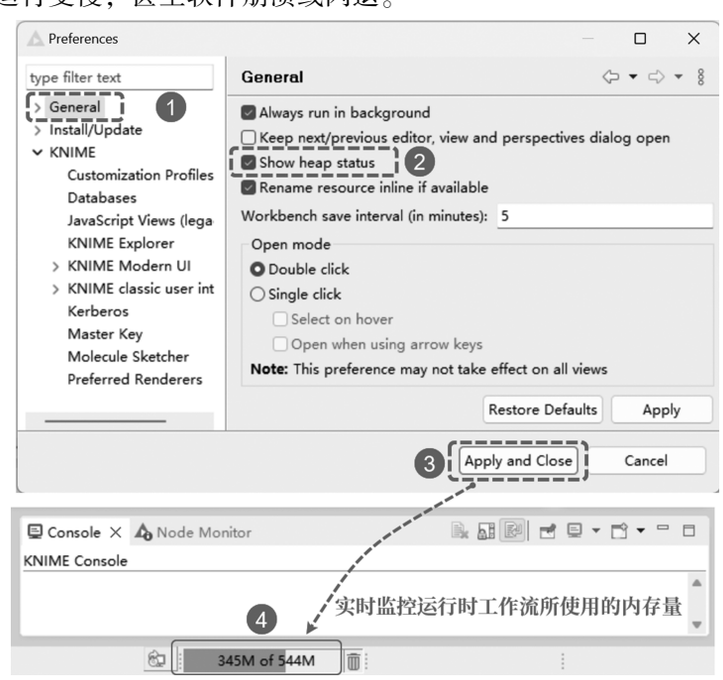


图 1-11　当前内存使用量显示设置（经典界面）

1.5　KNIME 扩展安装及卸载

KNIME 软件本身自带常用的各种节点，若这些节点不能满足自己的特殊需求，则可以选择在 KNIME 软件中安装丰富的扩展（Extensions）。一个扩展一般包含多个节点（Node）。这些扩展（插

件）从来源来说，既有来自 KNIME 官网的，又有来自第三方平台的。例如，NodePit 网站提供了额外的 KNIME 扩展，以及工作流和公开分享的组件（类似 KNIME Community Hub 功能）等。

本节主要介绍在线安装 KNIME 官方扩展的三种方式、安装非 KNIME 官方扩展的方式，以及如何删除扩展。以下截图主要以经典界面安装为例，现代界面安装类似，不再赘述。

1.5.1 在软件菜单中安装

在 KNIME 软件中，单击菜单 File→Install KNIME extensions，在弹出的新窗口中，通过输入关键词，搜索扩展。如希望安装支持 Python 编程语言的节点，则搜索关键词 Python，如图 1-12 所示，勾选 KNIME & Extensions 下的 KNIME Python Integration，单击 Next 按钮，继续后续各步操作。该扩展安装之后，在 KNIME 中 Python Script 节点中可直接编写代码。

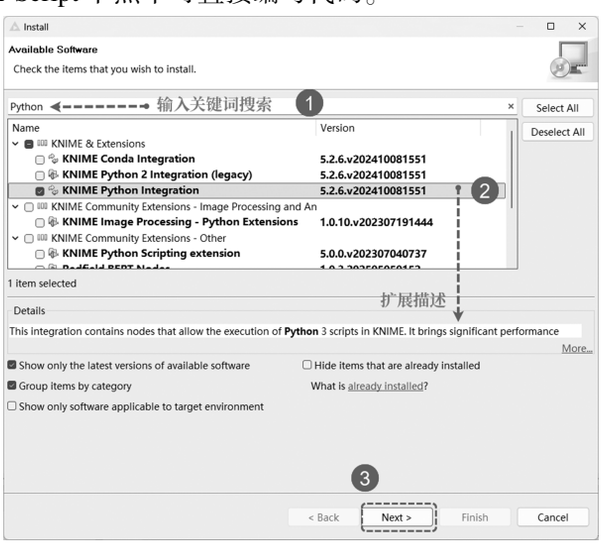


图 1-12 KNIME 安装扩展界面

1.5.2 KNIME Hub 拖曳安装

KNIME Hub 拖曳安装方式是将页面切换至 KNIME Hub 视图，或通过浏览器打开 KNIME Hub 官网（https://hub.knime.com/），搜索需要安装的扩展节点名称；找到该节点后，拖动左下角黄色图标至工作流编辑区（Workflow Editor）即可，KNIME 将在后台进行安装，如图 1-13 所示。该安装方式更适合明确知道需要的扩展节点名称，但不知道该扩展节点所属扩展的名称（一个扩展一般包括多个节点名称）。例如，希望安装 Python Script 节点，但不知其所属扩展模块名称，或扩展列表中有多个接近的扩展名称而又无法被准确辨认时，使用此方法更方便。通过拖曳安装某一个节点，将会完成安装该节点所属的整个扩展。



图 1-13 利用 KNIME Hub 拖曳节点完成扩展安装

1.5.3　导入工作流时安装

当导入他人分享的工作流文件，或从 KNIME Hub 上拖曳工作流至工作流编辑区时，由于本地没有安装相应的扩展，此时工作流中的小方块节点（Node）会出现红色叉号，KNIME 将自动分析该节点需要安装的扩展，并提示安装，这时单击 Next 按钮，继续完成安装即可，如图 1-14 所示。

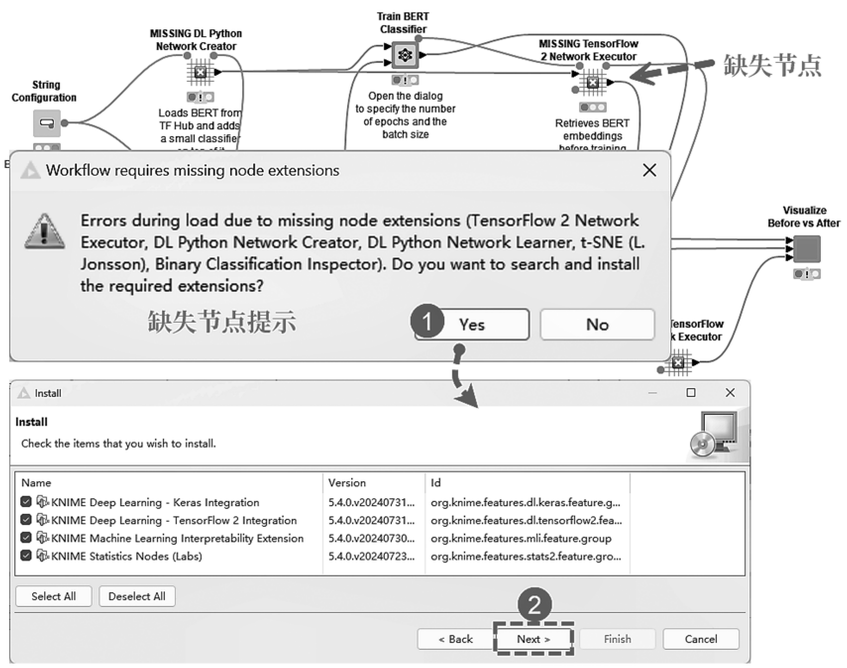


图 1-14　导入工作流时自动安装缺失节点对应的扩展

1.5.4　利用 NodePit 安装

NodePit 是一个与 KNIME 相关的独立在线平台，旨在为 KNIME 用户提供资源、支持和社区互动。具体来说，NodePit 是一个专注于 KNIME 节点（Nodes）和工作流（Workflows）的市场和社区平台，KNIME 用户可以在这里分享、查找和下载各种 KNIME 节点和工作流。NodePit 作为一个补充资源，通过提供额外的扩展，进一步增强了 KNIME 的实用性。

本节主要利用 NodePit 网站提供的丰富扩展，展示如何在 KNIME 中安装其提供的扩展。以安装 Palladian 扩展为例（本书第 12 章将利用该扩展的正则等节点），主要分为以下两个步骤。

第一步：在偏好设置 Preferences 窗口中设置可用更新站点，如图 1-15 所示。

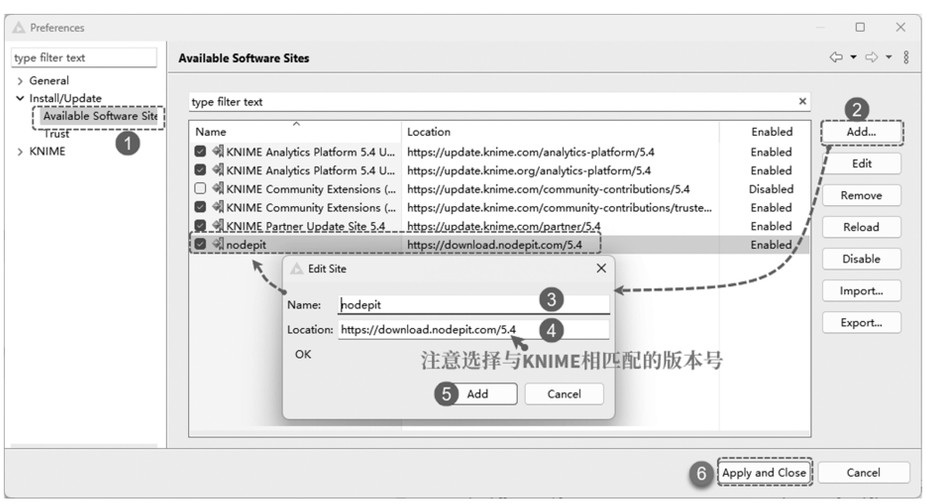


图 1-15　Preferences 中设置可用更新站点

第二步：在 KNIME 软件中，单击菜单 File→Install KNIME extensions，在安装扩展窗口中勾选带有“Palladian”关键字的各个复选框，然后单击 Next 按钮，如图 1-16 所示。后续各步骤按照新窗口的提示一步步操作。该扩展安装成功后，在节点库中将会新增一个模块 NodePit 及子模块 Palladian，其下有诸多新增节点，如本书中使用的正则节点 Regex Extractor 等。

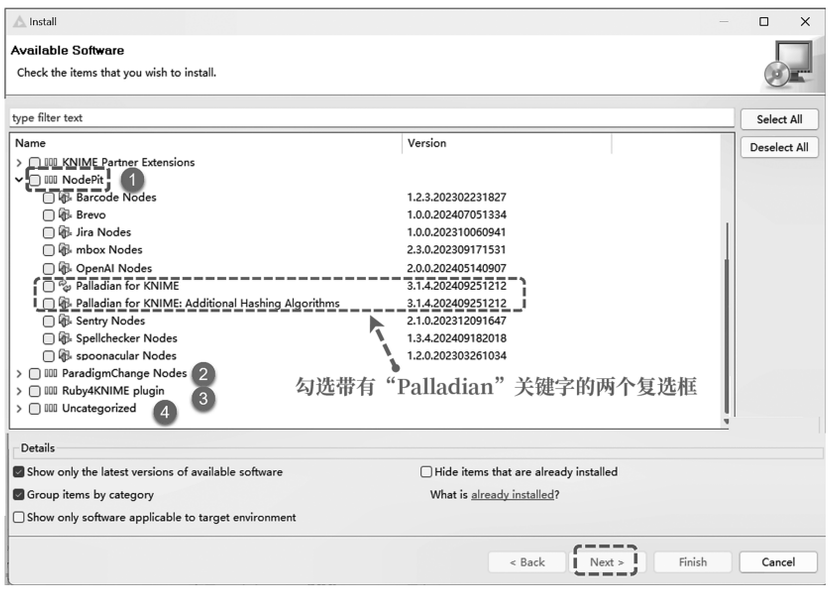


图 1-16　安装 NodePit 提供的 Palladian 扩展

注意，若需要安装 NodePit 其他扩展，也是同样的方法。该网站提供的各个扩展如图 1-17 所示，这些扩展全包含在图 1-16 所示的 NodePit、ParadigmChange Nodes、Ruby4KNIME plugin、Uncategorized 四个组别之中。

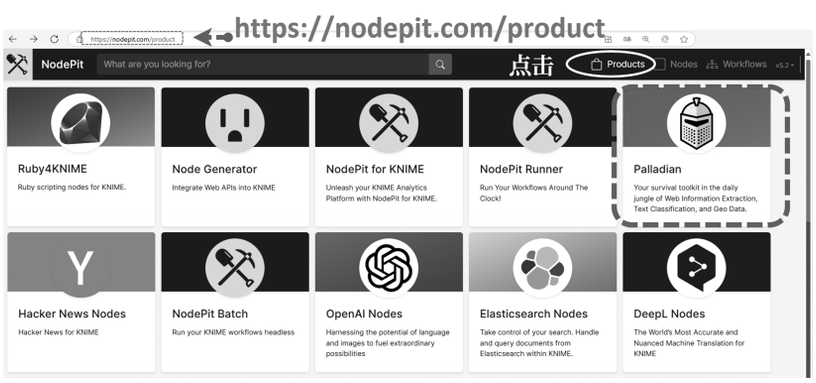


图 1-17　NodePit 网站提供的各扩展列表

四种安装 KNIME 扩展的方式都属于在线安装，若电脑处于隔离的环境中，或出于安全原因不能对外联网，也可以到 KNIME 官网下载扩展包压缩文件，通过离线方式安装。这属于安装扩展的特殊情况，操作步骤总体与图 1-15 和图 1-16 类似，请参考本书附带视频讲解。

1.5.5　删除扩展

删除扩展操作类似于上面的安装界面，单击菜单 File→Install KNIME extensions，再单击 already installed 选项，获取所有已安装的扩展列表，如图 1-18 所示；在列表中选择需要卸载的扩展，单击 Uninstall 按钮完成卸载，如图 1-19 所示。

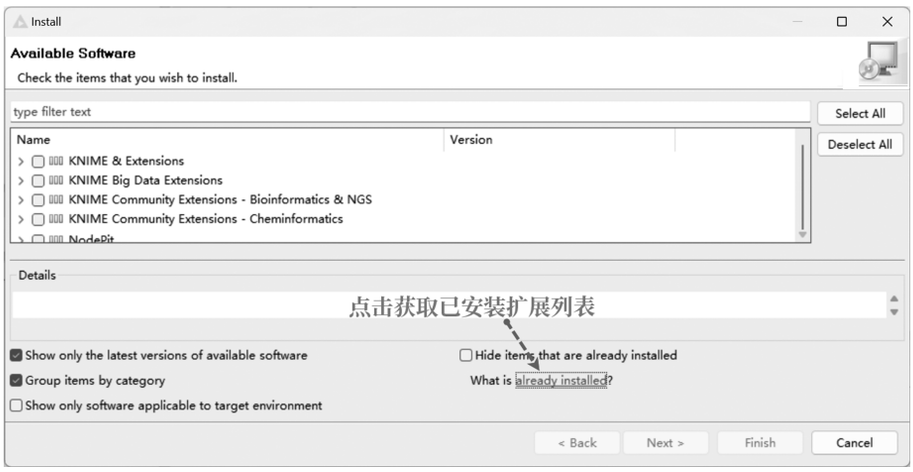


图 1-18　查看已安装扩展列表

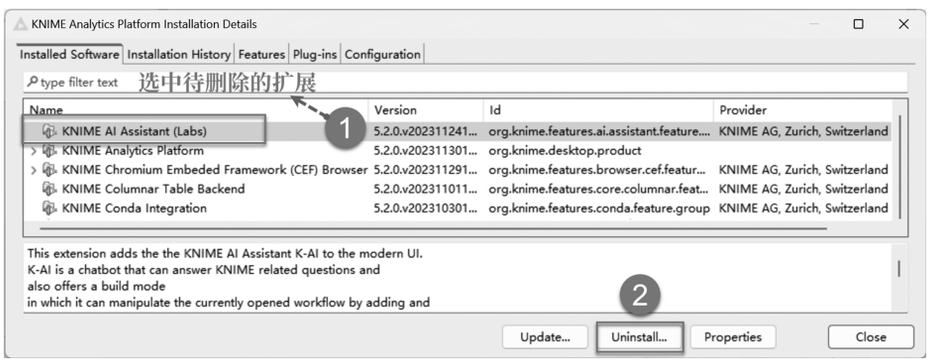


图 1-19　删除扩展

1.6　KNIME 两种界面的比较

1.6.1　经典界面

经典界面布局如图 1-20 所示，每个视图名称及功能概括如下。建议初学者快速浏览本节，了解各视图的作用和功能，以便有一个整体性的认识，个别不理解的内容，可等有一些基础后再学习。经典界面下重点要掌握好的视图编号为 1、3、4、6。

视图 1：KNIME Explorer，资源管理器。该视图默认主要分为三个部分。第一个是 My-KNIME-Hub，此处需要以注册的 KNIME 账号连接网络后进入。默认将创建一个同样注册账号名的空间中，可将工作流 Workflow 或组件 Component 存储在个人私有空间，或发布至公开可见的公共空间中，其他人可通过 KNIME Hub 搜索到。第二个是 EXAMPLES 工作流示例库，需要在线使用，是很好的学习资源库。第三个是 Local Workspace，即首次安装 KNIME 后启动时设置的文件保存路径。其下包含了部分示例工作流，是 EXAMPLES 中的一小部分，无须网络连接即可在本地查看，在此目录下可新建 KNIME 工作流和 KNIME 工作流组。

视图 2：Workflow Coach。该视图根据社区节点、本地使用节点的频率和各节点间的关联性进行工作。当在工作流中选中一个节点后，该视图将推荐下一个可能使用的节点。注意，该视图在现代界面中已消失，主要通过 KNIME 提供的 K-AI 智能工具替代。在现代界面中，基于当前节点，连接下一个空节点连线时将出现一个可能需要使用的下拉列表节点，该点对于有经验的用户来说会更加方便，稍后在现代界面详解。

视图 3：Node Repository，节点库。其下每一个节点都实现了一个特定的功能，一个工作流是由一

个个节点有机连接在一起实现的。此处是学习的重点，每个节点可类比实现某个功能的函数。在节点库中，可在右上角输入框中输入关键词，并基于关键词进行精准搜索，或者模糊搜索。节点库默认有13个模块，见表1-1所列。

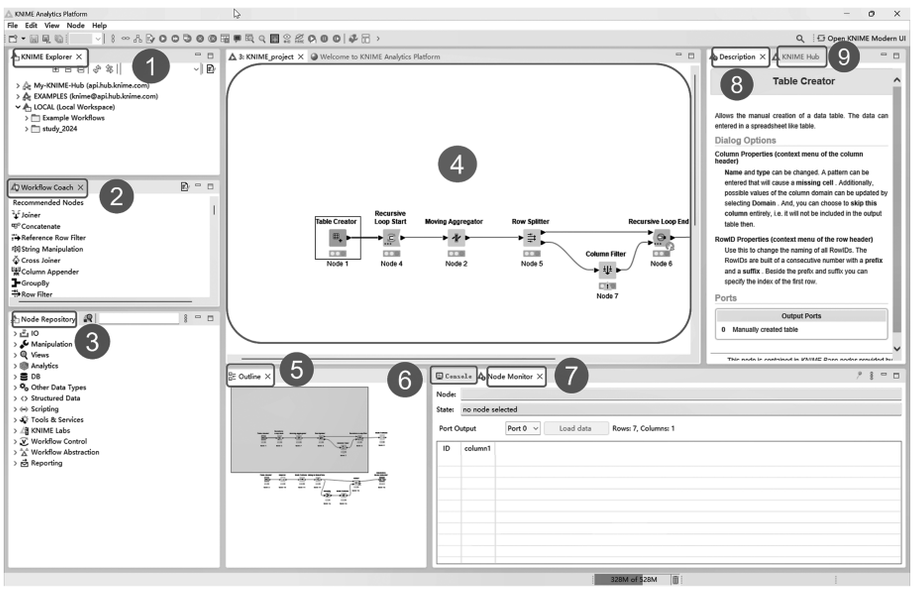


图1-20　经典界面默认视图

表1-1　经典界面各视图功能简介

编号	模块名	说明
1	IO	Input Output，输入或输出模块，里面包含各种读入类和写入类节点，以及与文件相关的其他操作，如常见的Excel Reader节点
2	DB	DataBase，支持各种常见数据库，如MySQL、Oracle、PostgreSQL等，通过直连数据库，进行更灵活的数据操作
3	Other Data Types	主要针对日期时间类型的各种操作节点，如日期作差、平移、日期格式转换、基于日期过滤等
4	Manipulation	包括行列过滤、数据清洗、格式转换、数据汇总、合并等操作，是数据处理和分析的最重要模块
5	Views	数据可视化，支持散点图、直方图、条形图、雷达图、多层环状太阳图等，也可基于ECharts实现各种图形，可在格式化报告中使用这些图表
6	Analytic	传统统计分析及机器学习各种模型，如回归、决策时、随机森林、神经网络、支持向量机等
7	Structure Data	主要针对JSON和XML格式数据的各种处理，如基于Xpath解析网页并提取数据
8	Scripting	默认仅有Java，若安装扩展（如Python、R等编程语言），它们则会出现在此处
9	Tool & Services	与网页相关的一些服务，不常用，初学者可忽略
10	KNIME Labs	实验室节点，若发展成熟，将加入正式版本的对应模块中，或者移除，此处的Column Expressions节点很有用
11	Workflow Control	对工作流的控制，如流变量、分支、循环、错误处理等，是进阶提升的必备模块

续表

编号	模块名	说明
12	Workflow Abstraction	提供自定义图形化界面节点，将主要参数设置在一个图形化界面中，实现更灵活的交互界面操作
13	Reporting	基于分析后的数据和图表等信息，整合至 PDF 或 HTML 报告中，实现从数据导入、处理、分析，到报告自动生成的一站式处理。该模块下的 BIRT 报告节点操作过于复杂，建议使用 KNIME 报告类节点，第 13 章将单独展开讲解

视图 4：Workflow Editor，工作流编辑区。将节点库的节点拖至该视图的编辑区，再将节点进行连接，即为工作流。该工作流保存路径为首次启动 KNIME 时设置的目录，该目录在本机的名称即为 Local Workspace，简称 Local。

视图 5：Outline，工作流缩略图。当整个工作流节点较多，查看工作流不方便时，可通过此视图实现对整个工作流的快速移动和浏览。

视图 6：Console，KNIME 控制台。当工作流运行时，对运行过程进行记录反馈，若工作流有错误或节点配置不正确，该视图将有相应的提示信息，对应节点调试非常重要。在 5.4 版本现代界面中，可点击左侧控制面板的 Monitor，查看工作流预警或报错信息。

视图 7：Node Monitor，节点监控。选中节点，即可查看该节点相关的数据。现代界面对其做了更多的改进，对于数据查看、排序、搜索等更加友好，该点值得称赞。经典界面下该视图与 Console 并列排列，可通过单击标签页切换。

视图 8：Description，节点帮助文档。选中节点，即可看到该节点的功能概述、界面参数的详细解释，甚至学习资料的一些链接，新手务必把常用的节点文档认真看一遍。若遇到不知道节点参数如何设置的问题，可随时查看该节点帮助文档。

视图 9：KNIME Community Hub（简称“KNIME Hub”），KNIME 提供的公开分享库，主要有四种类型，分别为工作流（Workflows）、节点（Nodes）、组件（Components）、扩展（Extensions），这些资源是宝贵的在线学习资源。可通过此界面便捷安装扩展、下载组件、工作流等。在 KNIME 现代界面中，没有该视图，可通过浏览器打开 KNIME Hub 网页使用，如图 1-21 所示。

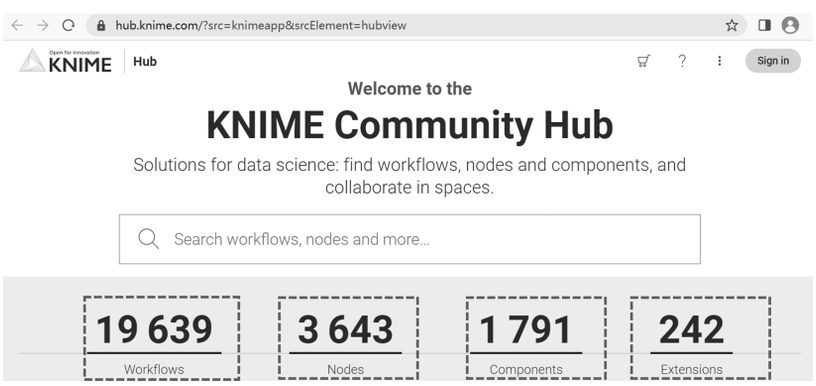


图 1-21　KNIME Community Hub 资源

如果以上各个视图窗口最小化或关闭后找不到了，或者默认布局被打乱，可通过点击菜单 View→Reset Perspective 恢复默认视图。

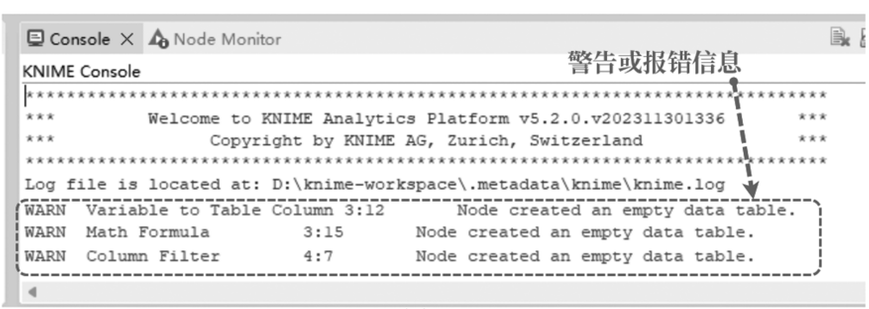
1.6.2　现代界面

KNIME 默认启动界面为现代界面，现代界面和经典界面可随时切换。若当前界面为经典界面，可单击经典界面右上角按钮 Open KNIME Modern UI，一键切换至现代界面，如图 1-22 所示。现代界面相对经典界面做了减法，主要变化总结如下。

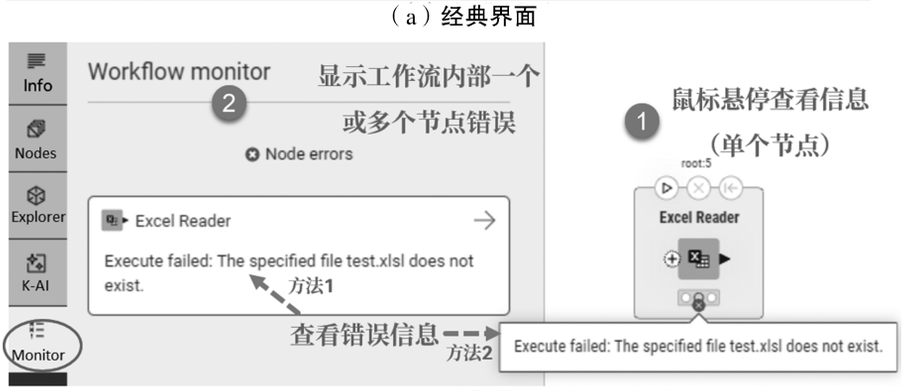


图 1-22　现代界面 Modern UI 各视图

（1）删除 Outline 视图、Console 视图和工具栏。现代界面移除了 Outline 视图，该功能可通过“Ctrl+鼠标滚轮”实现窗口的工作流编辑区的放大缩小。通过长按“Space+光标移动”实现工作流的浏览。此外，现代界面还移除了控制台 Console 视图，重构了该功能。在经典界面和现代界面中，查看工作流相关的警告或错误提示如图 1-23 所示。在 5.4 版本现代界面中，鼠标悬停至报错节点时可查看报错信息，弥补缺失功能；或者点击左侧面板中的 Monitor 切换至 Workflow Monitor 视图（功能同经典界面中的 Console 视图），可查看多个节点的预警或报错信息。



（a）经典界面



（b）现代界面

图 1-23　经典界面与现代界面查看错误提示的方式

（2）新增 KNIME AI Assistant 视图。该智能助手是现代界面新增功能，可通过英文提问，得到对解决问题的建议等，如图 1-24 所示。5.4 及后续版本默认已安装 K-AI，之前版本需要另行安装扩展。使用该智能助手时，需要先登录个人的 KNIME 账号。K-AI 主要有问答模式和构建模式，协助用户构建工作流。此外，在一些编程类节点中，如 Python Script、Generic ECharts View 等节点也提供 K-AI 智

能问答，辅助生成代码。

（3）优化 Node Monitor 视图。在现代界面中查看节点数据时，可实现更灵活的节点排序、查看、搜索，以及数据的常用统计值，浏览数据表、流变量也更加方便。

（4）KNIME Explorer 和 Node Repository 视图，分别如图 1-25 和图 1-26 所示。在现代界面资源管理器中，若存在多个工作流组 A 和 B，双击才能进入 A 工作流组，进而看到 A 内部的工作流组，此时就看不到 B 工作流组的内容。可见，不同工作流组下的工作流浏览不太方便。在节点库视图中，缺少节点是否支持流式处理的信息。而在经典界面中，浏览工作流和节点非常方便。此外，现代界面中，KNIME Explorer 视图移除了官方的在线工作流示例库，如需要浏览示例，应切换至经典界面。

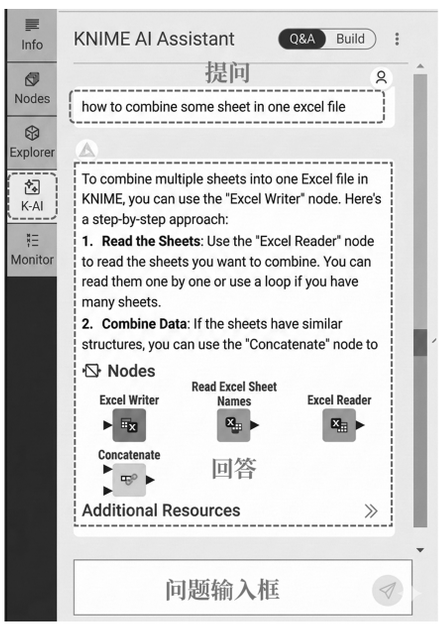


图 1-24　KNIME AI Assistant 视图

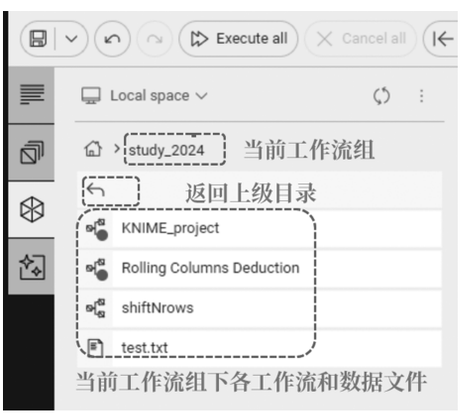
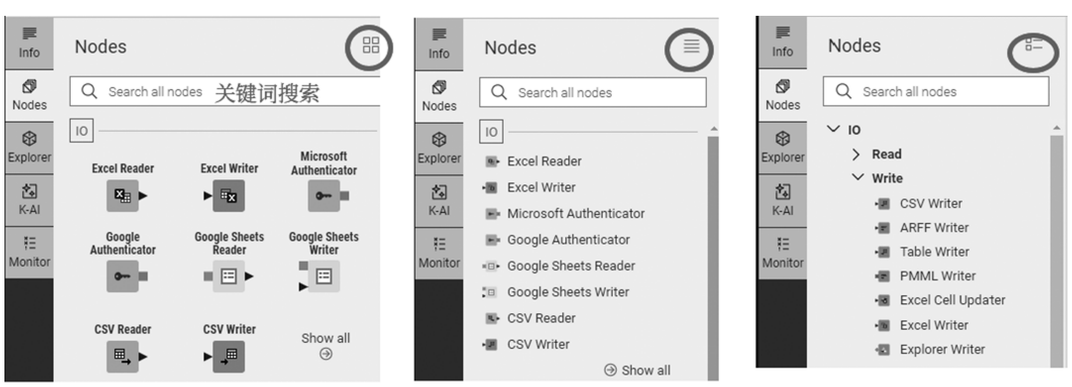


图 1-25　现代界面中的 Explorer 界面



（a）平铺　　（b）列表　　（c）树状

图 1-26　现代界面节点库的三种不同视图

（5）KNIME Editor 视图中增加了更多便捷操作，熟悉 KNIME 的用户可快速在工作流中添加节点。例如，从当前节点直接连线下一个节点，连线时可从下拉菜单中选取节点，或根据关键词搜索节点，如图 1-27 所示。此外，当鼠标悬浮于节点上方时，节点自动出现两个红色流变量的输入和输出端口，以及配置、运行等四个小按钮。因此对于配置和执行工作流来说，现代界面比经典界面更方便。另外，现代界面工作流注释框内的文本新增了支持项目编号、列表等更多格式，节点之间的拖动也变得更加流畅，可实时动态显示，没有经典界面的拖动节点的卡顿感。这些改进可圈可点，相比在经典界面下操作方便了很多。



图 1-27　现代界面下工作流编辑区新增快速选择节点功能

1.6.3　两种界面优缺点比较

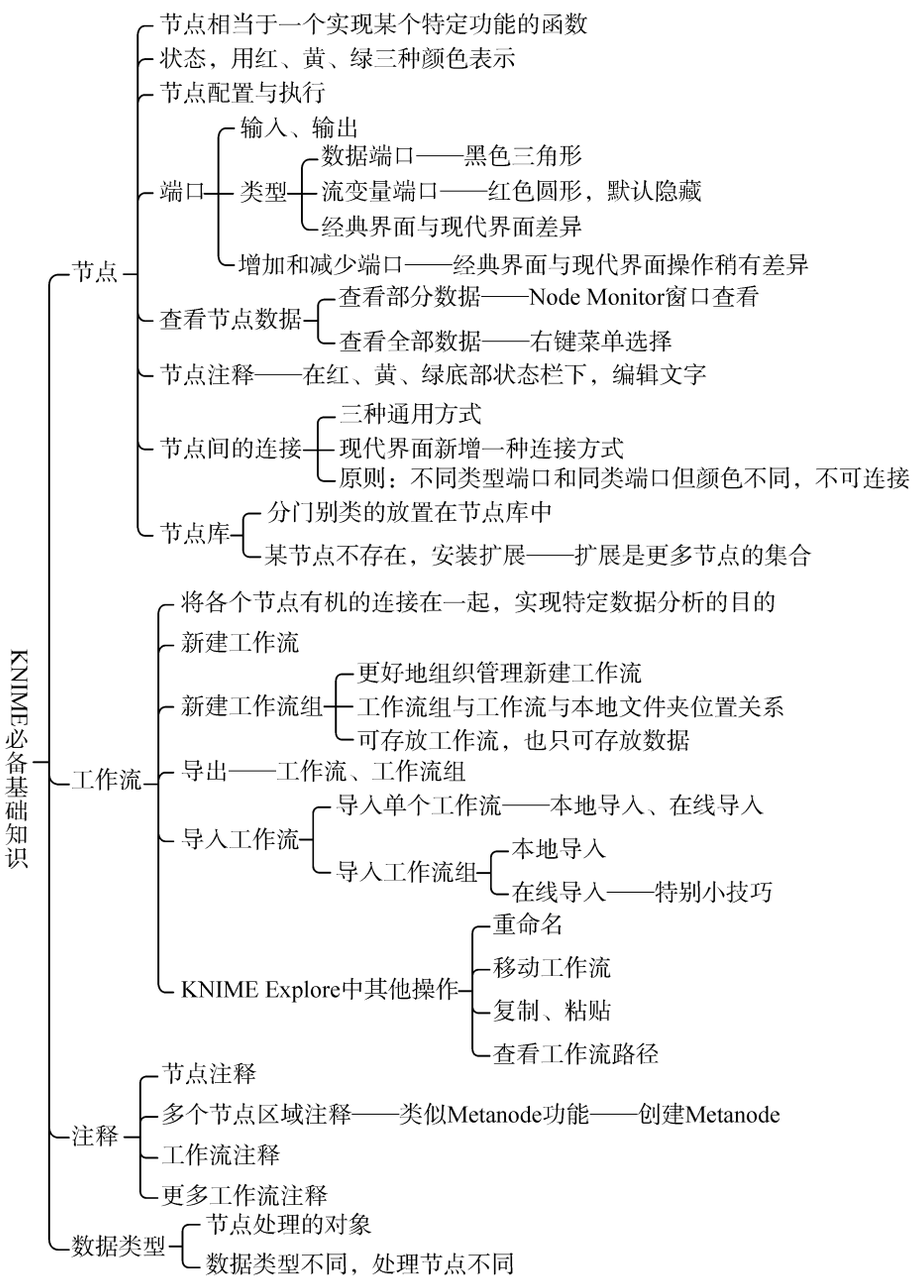
KNIME 新版本默认的现代界面在工作流浏览、在线安装扩展、工作流调试等方面存在诸多不便，现将经典界面和现代界面的比较列于表 1-2 中。对于初学者来说，强烈建议先学习经典界面下的操作，待熟悉之后可随时切换至现代界面。当然，现代界面也是有很多可圈可点的改进和优点，如数据查看、添加节点、智能助手、工作流注释等。而且，需要注意的是在使用过程中现代界面和经典界面随时可以切换，不必二选一。

表 1-2　经典界面与现代界面视图比较

编号	视图名称	经典界面	现代界面
1	KNIME Explorer	支持树状层次结构，随意展开工作流组工作流，查看比较方便	缺少层次结构，浏览不同工作流组的工作流不方便
2	Workflow Coach	根据使用频率等推荐当前节点可能使用的下一个节点	删除了该视图，新增 K-AI 助手替代
3	Node Repository	树状结构显示各个模块的节点类别，可以查看节点是否支持流式处理（Streamable）	5.4 版本及之后的版本支持树状结构浏览各节点，不能查看节点是否支持流式处理
4	Workflow Editor	传统式拖曳或搜索方式新增节点，通过双击或右击选择 Configure 对节点进行配置	改进了注释，支持更多格式排版，以及新增连接下一个节点的便利操作。选中节点浮动出现配置、执行等小按钮，直接操作，操作更加方便
5	Outline	工作流缩略图快速导航	5.x 版本删除了该视图，借助“space+鼠标移动”放大缩小工作流快捷键来实现导航。5.6 版本增加该功能
6	Console	支持不同类型的日志记录和显示，有 Debug、INFO、WARN、ERROR 四种模式。该视图对工作流调试极为重要，查看错误信息更方便	在 5.x 版本中删除该控制台视图。在 5.4 版本中，鼠标悬停至报错节点或点击左侧 Monitor 可查看节点错误信息
7	Node Monitor	显示当前节点的数据，不支持排序、搜索等	显示当前节点数据和流变量，数据概览和查看更方便
8	Description	在右边显示当前选中的节点文档	选中工作流的当前节点时，左边显示节点文档，仅视图位置变化而已
9	KNIME Hub	支持从 KNIME Hub 界面内向工作流拖曳节点安装节点或者扩展，较为方便	删除了该视图，需要从浏览器网页端打开 KNIME Hub，稍有不便
10	工具栏	节点位置对齐调整，执行等功能	悬停节点会在上面出现配置、执行等的小图标，操作更便利。右键选择水平或垂直方向对齐各节点等

第 2 章　KNIME 必备基础知识

本章主要介绍 KNIME 的必备基础知识，包括 Node 节点、Workflow 工作流、新建工作流和工作流组、导出和导入工作流等。此外，还介绍了常见数据类型的转换，以便后续根据不同的数据类型选择相应的节点，为初学者掌握 KNIME 奠定基础。本章主要知识点总结如下。



2.1 节点（Node）

节点（Node）是实现某个功能的一个环节，或视为一个具有输入、输出和图形配置窗口的封装函数，节点库中每个方块为一个节点。节点需要设置参数，其底部显示节点状态，左右两侧一般有输入端口和输出端口，通过连线将各个节点连接起来，实现一系列的数据操作。

为了让读者对本章有个整体的了解，以下通过一个截图展示各个关键概念，如图 2-1 所示。该图涉及节点、端口、节点状态、工作流、工作流管理等基本知识。简言之，节点就是处理数据的一个环节，通过节点之间的连线组合成一个数据处理流程，即工作流，保存在左侧的 KNIME Explorer 资源管理器中。下面，将对图中涉及的各个知识点逐一讲解。

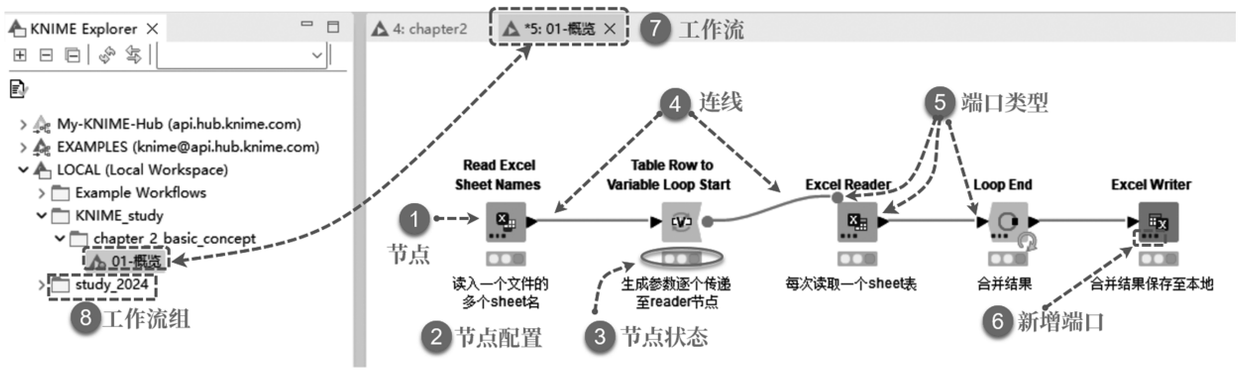


图 2-1 本章关键概念知识点概览

2.1.1 节点状态

KNIME 通过创建工作流来实现数据处理和分析，节点为工作流的最小单位。在经典界面下，它们在 KNIME 节点库中按照不同模块和层级存放，根据需要可将各个小方块节点拖动到工作流编辑区中进行工作。节点的三种状态由节点底部的颜色表示，有红、黄、绿三种颜色，类似交通指示灯，如图 2-2 所示。有时，在该节点状态底部有一个带有感叹号的黄色三角形或红色叉号，分别代表对该节点的警告或错误信息，在底部的 Console 窗口中可以查看具体日志信息。

- 红色：节点未配置。正确配置后，第一个红色指示灯将跳转至第二个黄色指示灯。
- 黄色：已正确配置或无须配置，等待执行。
- 绿色：节点执行成功。若执行失败，则显示叉号。

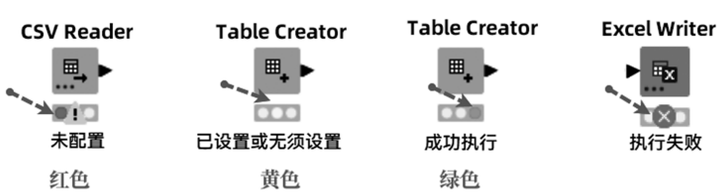


图 2-2 节点不同状态的颜色

2.1.2 节点配置和执行

在节点执行前，一般需要先设置相应的参数或选项，配置正确后，才可以运行。在经典界面下，双击节点或从右击菜单中选择 Configure，打开该节点的配置窗口，如图 2-3 所示。在现代界面中，鼠标移动到该节点上方，可看到节点上方出现四个小图标，选择小齿轮图标即可进行配置。

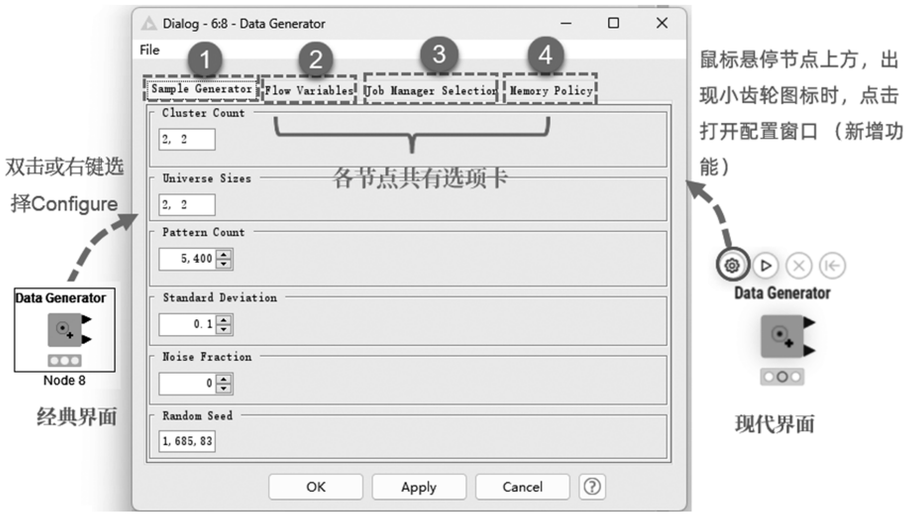


图 2-3　节点设置通用窗口示例

除了图 2-3 中打开节点配置窗口的三种方法外，KNIME 5.5 版本增加了第四种方法（现代界面中默认），即内嵌于 KNIME 右侧界面中，不再重新打开一个新的节点窗口，如图 2-4 所示。若不习惯该方式，可以在 Preferences→KNIME Modern UI 中设置，在节点配置对话模式中选择 Open in new window（默认为 Embedded inside application），这样就恢复至双击打开单独的节点配置窗口了。注意，不是所有节点都支持这种内嵌式节点配置界面。

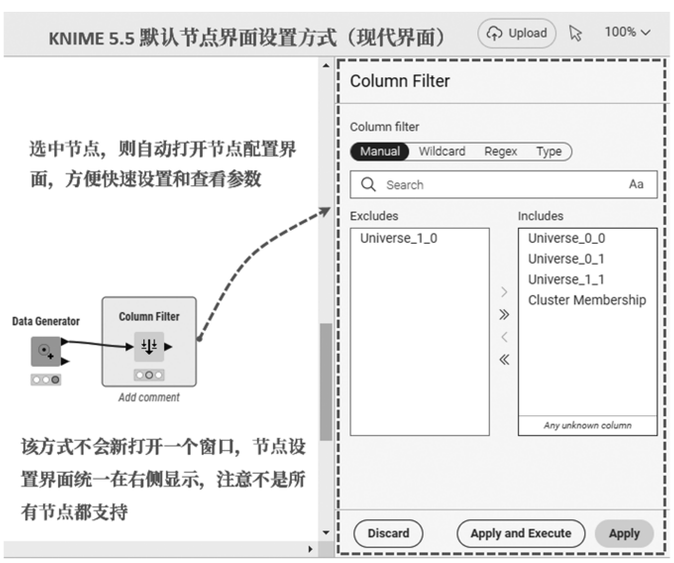


图 2-4　新增内嵌式节点界面设置（现代界面）

节点窗口一般至少有四个标签页（也称“选项卡”，本书统一使用标签页），设置参数的标签有一个或多个，一般需要进行设置，部分节点已默认设置参数。另外，最右侧的三个选项卡为各节点通用标签页，其中流变量的配置和使用涉及 Flow Variable 标签页，在第 9 章中再深入讲解。Job Manager Selection 和 Memory Policy 标签页，一般使用默认即可，高级操作在涉及时再展开（本书第 16 章有涉及此标签页的设置）。个别节点的流变量设置并不在 Flow Variable 标签页，需要在节点右击菜单才能找到，第 9 章流变量章节再展开。

节点配置正确后，其底部状态显示黄色时即可执行。选中节点后，右击选择 Execute，或者在经典界面下的工具栏中选择执行按钮。注意，现代界面下无工具栏，直接单击节点上方浮动的三角形执行按钮，如图 2-5 所示。

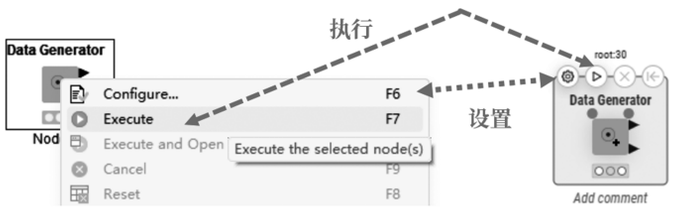
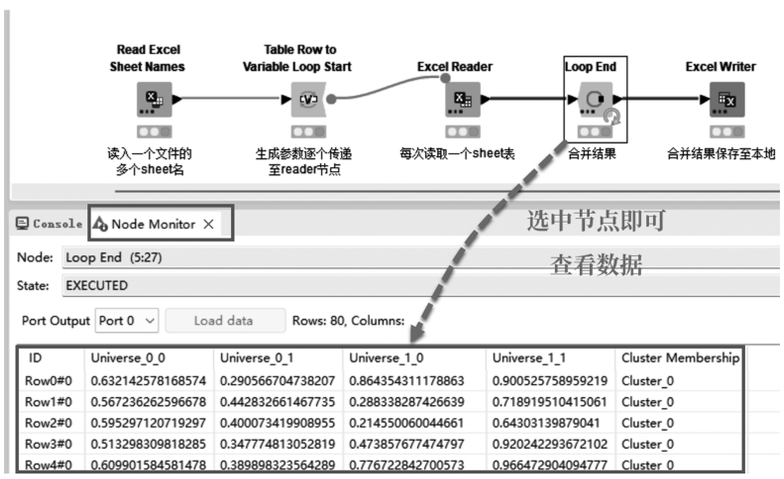


图 2-5　在经典界面与现代界面下设置及执行

2.1.3　节点数据查看

节点成功执行后即可查看数据。查看数据有两种方式。

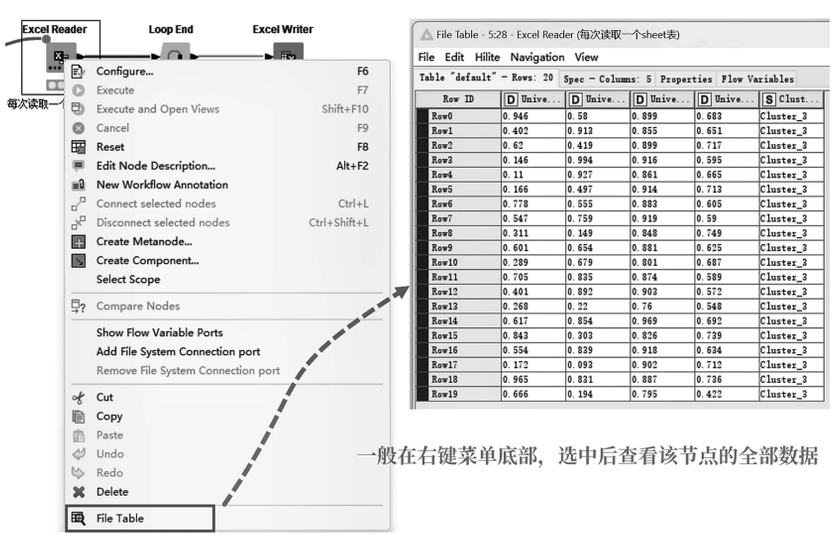
（1）在经典界面下，可快速查看部分数据，如图 2-6 所示。在工作流编辑区选中已成功执行过的节点，在界面右下角 Node Monitor 选项卡下将显示该节点输出的数据表。注意，由于该界面空间有限，只显示部分数据，使用该方式可快速浏览各个节点的数据。



ID	Universe_0_0	Universe_0_1	Universe_1_0	Universe_1_1	Cluster Membership
Row0#0	0.632142578168574	0.290566704738207	0.864354311178863	0.900525758959219	Cluster_0
Row1#0	0.567236262596678	0.442832661467735	0.288338287426639	0.718919510415061	Cluster_0
Row2#0	0.595297120719297	0.400073419908955	0.214550060044661	0.64303139879041	Cluster_0
Row3#0	0.513298309818285	0.347774813052819	0.473857677474797	0.920242293672102	Cluster_0
Row4#0	0.609901584581478	0.389898323564289	0.776722842700573	0.966472904094777	Cluster_0

图 2-6　选中节点后在 Node Monitor 界面下查看数据

（2）详细查看全部数据。工作流编辑区选中已成功执行过的节点，右击最底部的菜单选项，如图 2-7 所示。使用该方式可以对整个表进行全面的数据查看。同时，可以对数据表进行临时的数据排序和搜索，显示小数为百分比，条形图显示数据大小，也可任意拖动列等。这些操作仅限于浏览数据，不会对原数据产生影响，如将最后一列拖动至第一列，原始数据不发生变化。



Row ID	Unive...	Unive...	Unive...	Unive...	Clust...
Row0	0.946	0.58	0.899	0.683	Cluster_3
Row1	0.402	0.913	0.855	0.651	Cluster_3
Row2	0.62	0.419	0.899	0.717	Cluster_3
Row3	0.146	0.994	0.916	0.595	Cluster_3
Row4	0.11	0.927	0.861	0.665	Cluster_3
Row5	0.166	0.497	0.914	0.713	Cluster_3
Row6	0.778	0.555	0.883	0.605	Cluster_3
Row7	0.547	0.759	0.919	0.59	Cluster_3
Row8	0.311	0.149	0.848	0.749	Cluster_3
Row9	0.601	0.654	0.881	0.625	Cluster_3
Row10	0.289	0.679	0.801	0.687	Cluster_3
Row11	0.705	0.835	0.874	0.589	Cluster_3
Row12	0.401	0.892	0.903	0.572	Cluster_3
Row13	0.268	0.22	0.76	0.548	Cluster_3
Row14	0.617	0.854	0.969	0.692	Cluster_3
Row15	0.843	0.303	0.826	0.739	Cluster_3
Row16	0.554	0.839	0.918	0.634	Cluster_3
Row17	0.172	0.093	0.902	0.712	Cluster_3
Row18	0.965	0.831	0.887	0.736	Cluster_3
Row19	0.666	0.194	0.795	0.422	Cluster_3

图 2-7　使用右击菜单查看该节点执行后的详细数据

2.1.4　节点端口类型

节点端口类型一般有一个或多个输入或输出端口，按照节点端口传递的对象不同，可将节点端口分为三种，如图 2-8 所示。

- 黑色三角形：代表数据集（数据表），是最常用端口。
- 红色圆形：工作流变量，简称“流变量”，第 9 章再详述。
- 正方形端口：除上述数据和流变量之外的所有对象，如传递的可能是一个数据库连接参数、图像、模型、报告等。该类型种类较多，每种类型可通过颜色区分，不同颜色的方形端口之间一般不可以连接。

KNIME 中只有相同类型的端口才可以连接，即数据端口连接数据端口，流变量连接流变量。该规定强化了对端口的一致性管理，减少了人为错误。如图 2-8 所示，数据库操作节点中标号 3 和标号 4 都为方形端口，但节点端口颜色不一样，这表明两者对象不一致，不能连接。当然也存在个别例外，可参考本书第 13 章图 13-28 工作流中的报告组件节点与 Model Loop End，之所以能连接是因为后者输入端口兼容上游报告节点输出的结果。

在各类端口中，也存在实心与空心两种类型。如图 2-8 所示，图中标号 5 左侧为红色空心圆形，代表流变量端口。一般来说，空心三角和空心圆代表可以通过上一个节点获取数据或流变量，也可以通过本节点内部设置而无须上一个节点传递，即代表可选项，非必须连接点。

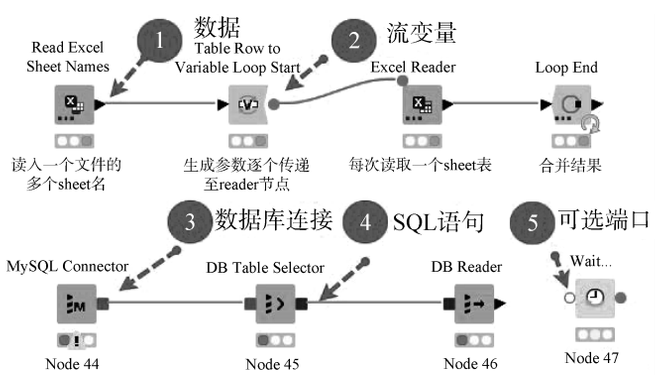


图 2-8　节点端口类型

2.1.5　节点连接

如图 2-9 所示，节点库中的节点一般通过拖动或者双击节点至编辑区，然后再将各个节点连接。在经典界面下，前一个节点与下一个节点相连接时，有如下三种方式。

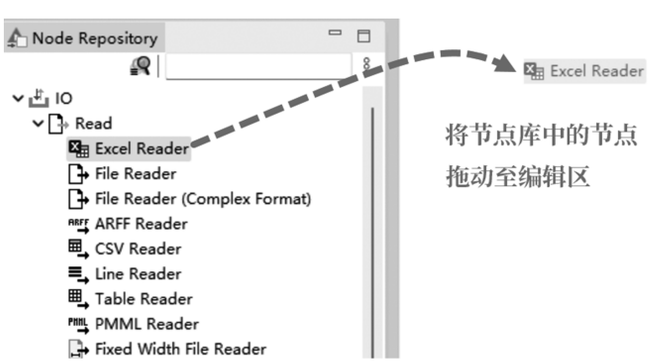


图 2-9　拖动节点库中的节点至工作流编辑区

第一种：在工作流编辑区内，鼠标选择左侧节点输出端口，按住左键直至连接下一个节点的输入端口再松开，如图 2-10 所示。

第二种：在工作流编辑区先选中 Excel Reader 节点，再在节点库中双击 Column Filter 节点，如图 2-11 所示。

第三种：在若想在已经连接的 A 和 B 节点之间插入 C 节点，可拖动 C 节点至 A 和 B 节点的连线上，此时连线颜色将由黑变红，松开即可将 C 节点插入 A 和 B 之间，如图 2-12 所示。

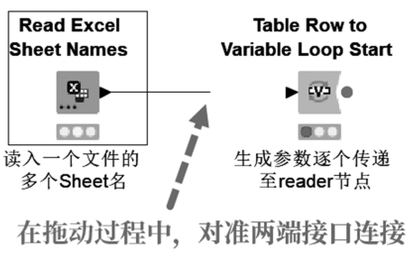


图 2-10　编辑区通过拖动左侧节点端口至右侧节点端口

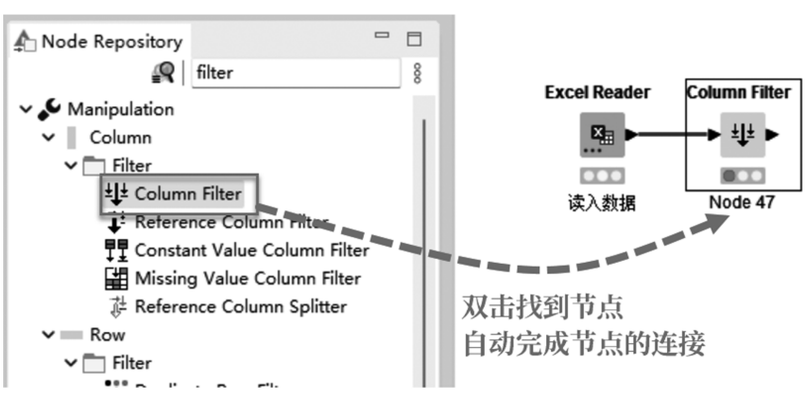


图 2-11　通过双击节点库中的节点自动连接

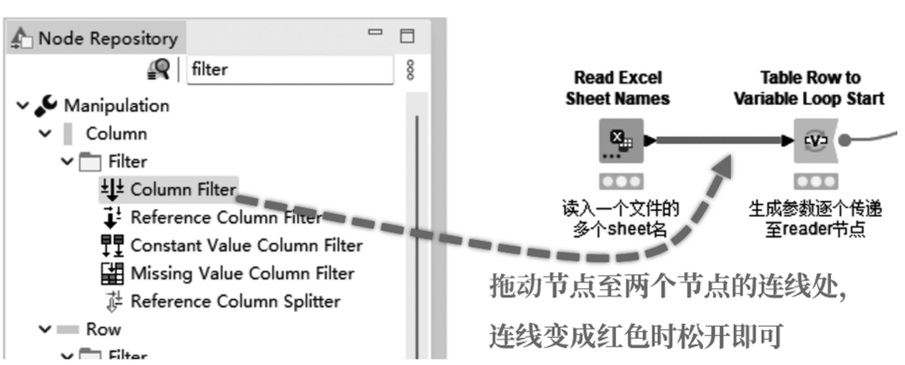


图 2-12　节点之间插入新节点

在现代界面下，除了上述三种方法外，多了另外一个节点间连接的便捷方式。在编辑区内选中一个节点端口，向右拖曳，将会跳出搜索栏，方便快速选择节点，或输入关键词搜索节点，如图 2-13 所示。该方法相比在节点库中搜索节点更加便捷，更适合熟悉节点名称且知道下一个要使用哪一个节点的用户。资深用户用此功能会更加方便，初学者推荐先多熟悉经典界面下各节点库中的节点名称、位置和所属模块。

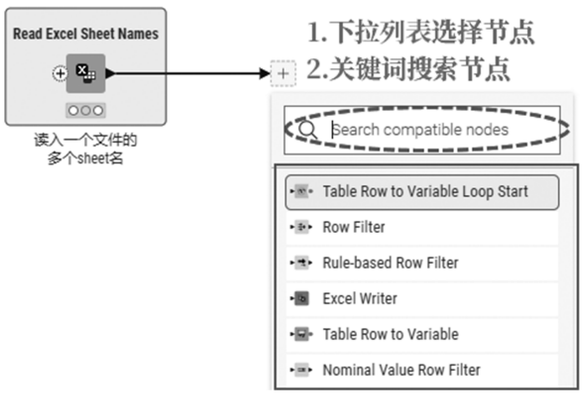


图 2-13　选择下一个节点并连接

2.1.6　节点端口增加和减少

部分节点的端口可以增加。例如，上下方向合并数据表的 Concatenate，可以从默认的两个输入端口增加到多个输入端口。由于 KNIME 存在两种不同界面，增加端口形式上稍微存在差别。

（1）经典界面增加端口的方法如图 2-14 所示。单击节点左下方三个点号处，将出现菜单，选择 Add input port 将会由两个端口，变更为三个端口。可以根据需要增加多个端口。其他类型的端口的增加也类似。在删除端口时，同样单击左下角三个点号处，选择 Remove input port。

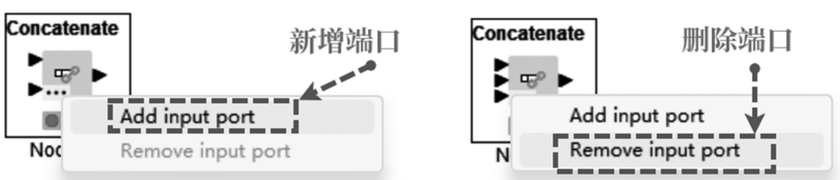


图 2-14　在经典界面下新增和删除端口

（2）现代界面增加端口的方法如图 2-15 所示。单击节点左下方加号即可新增节点，相对于经典界面新增端口操作更快捷。在删除端口时，选中端口，左边将出现一个垃圾桶小图标，单击该图标即可删除新增的端口。

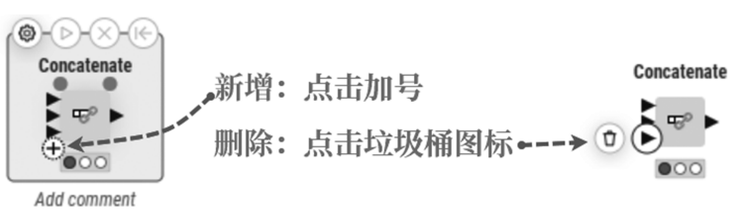


图 2-15　在现代界面下新增和删除端口

此外，还有一些端口需要在节点配置窗口勾选选项，才能增加端口，如图形类节点默认没有图形输出端口，除非勾选 Generate Image 选项。

2.2　工作流（Workflow）及工作流组（Workflow Group）

在工作流编辑区将一个个节点连接，实现一定的功能，这些节点构成了工作流。在 KNIME Explorer 资源管理器中，为了更好地组织各个工作流，如按照项目或日期等规则，可建立一个工作流组，以便更有条理地管理各个工作流。工作流组类似于集中管理各工作流文件而建立的文件夹。

对于工作流的操作主要涉及新建工作流，导入、导出工作流，工作流的复制、粘贴、重命名、移动等，本节内容皆为经典界面下的操作，现代界面下的操作稍有不同，不再对比说明。

2.2.1 新建工作流

单击软件左上角 KNIME Explorer，在 LOCAL 目录（首次启动软件时设置的本地保存位置）下选择浅蓝色文件夹图标的工作流组，右击新建一个工作流，输入工作流名称，单击 Finish 按钮后即可看到一个空白编辑区，如图 2-16 所示。此时即可拖动节点至工作流编辑区，构建工作流，最后保存工作流。



图 2-16 新建工作流

2.2.2 新建工作流组

新建工作流组类似新建工作流，单击 KNIME Explorer，在 LOCAL 目录下右击选择 New Workflow Group，类似图 2-16。值得注意的是，在 KNIME 中新建工作流组，KNIME Explorer 界面增加了一个浅蓝色的文件夹，对应本地位置多了一个文件夹，如图 2-17 所示。工作流组下面还可以新建其他工作流组。其中 LOCAL 代表本地保存的路径，一般为首次启动 KNIME 时设置的工作流保存路径。后面涉及读入数据文件路径时，可再深入查看该目录文件夹内工作流的结构。

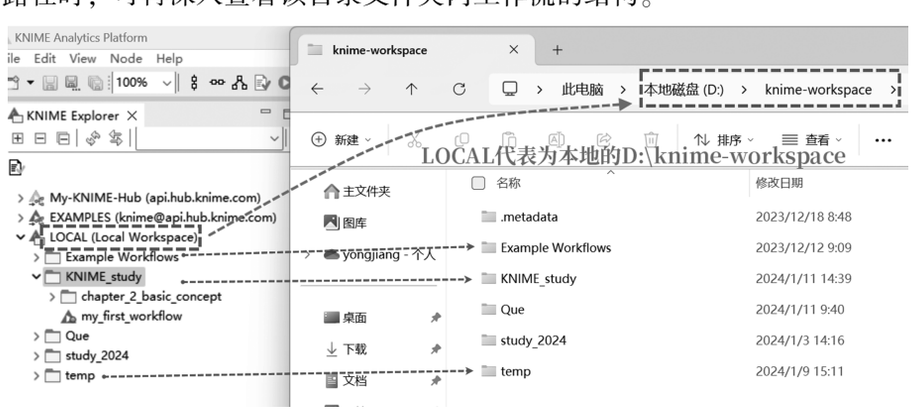


图 2-17 工作流组与本地目录对应关系

此外，工作流组不仅可以存放工作流，也可以存放数据、自定义组件（参见本书 15.2 节），还可在其下新建一个或多个子工作流组，如 data 文件夹。该文件夹仅存放数据文件，如 Basic 工作流组内各个工作流需要使用的数据，以便工作流读入文件时使用相对路径，如图 2-18 所示；并且在导出该工作流组时，也一并将其子文件夹下的数据附带导出，他人导入时也能直接运行涉及读入数据文件的工作流，稍后在读入文件时再展开。

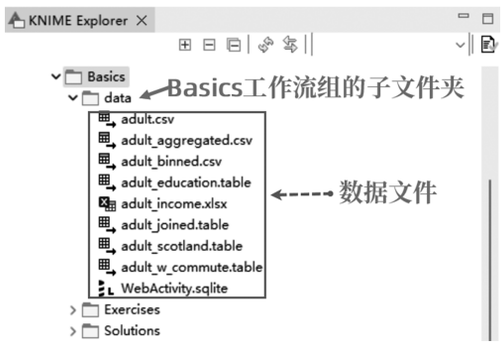


图 2-18　data 子文件夹及数据文件（子工作流组）

概括来说，在资源管理器中可保存工作流组、工作流、各数据文件、共享组件。其中，共享组件（Shared Component）可视为自定义节点，是从资源管理器中拖入工作流的，而不是从节点库中拖入的，自定义节点相关内容请参考第 15 章。

2.2.3　导出与导入工作流

KNIME 工作流有两种格式，后缀为. knwf 和. knar，分别是 KNIME Workflow 与 KNIME Archive 的缩写，对应单个工作流和工作流组。选中单个工作流，右击选择 Export KNIME Workflow，通过浏览本地路径导出工作流。单个工作流在导出时，后缀为. knwf，如图 2-19 所示。

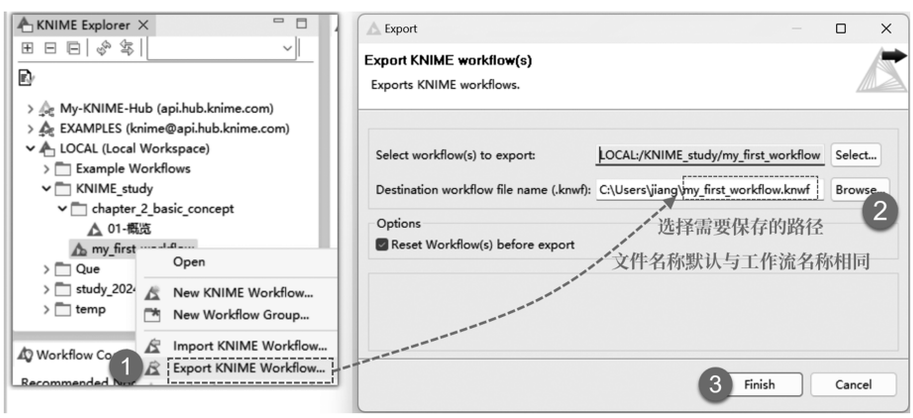


图 2-19　导出单个工作流

若需要导出某个工作流组下的多个工作流，选中该工作流组，右键选择 Export KNIME Workflow。由于工作流组存在多个工作流，可在导出前选择需要导出的工作流及子工作流组，导出后文件后缀为. knar，内部包括数据、子工作流组、工作流等，如图 2-20 所示。

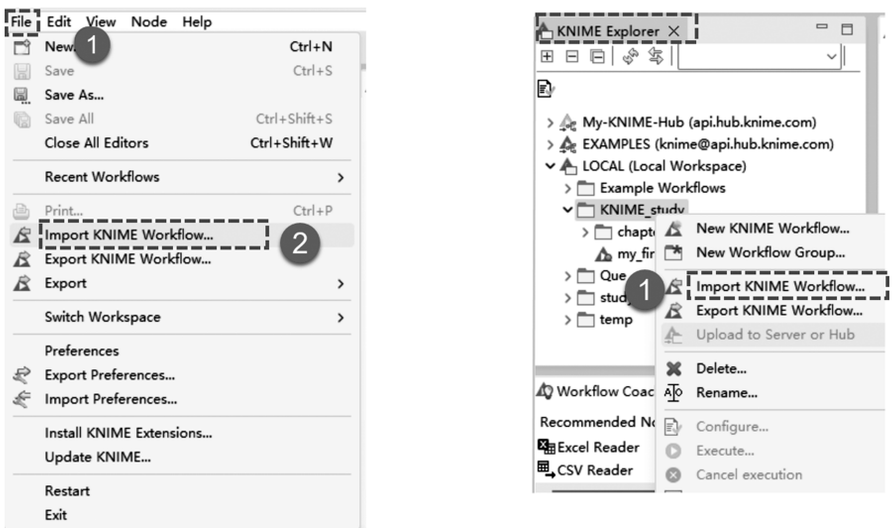


图 2-20　导出多个工作流

2.2.4 导入工作流和工作流组

由于导出的工作流分为两种后缀格式，即后缀为. knwf 的单个工作流和后缀为. knar 的工作流组。两者导入时操作类似，如图 2-21 所示。

导入工作流或工作流组文件时，可以通过点击 File→Import KNIME Workflow，或者在 KNIME Explorer 界面内，选择对应位置文件夹导入。其中在导入后缀为. knar 的文件时，可以选择导入该文件中的全部或部分工作流。



（a）方法1　　（b）方法2

图 2-21　导入工作流或工作流组

除了上述常规的导入方式外，KNIME 还提供了在线导入本地工作流的方式。即在经典界面的 KNIME Hub 窗口中，搜索需要的工作流，将右下角黄色方块小图标拖曳到工作流编辑区，将会从远程下载至本地，如图 2-22 所示。若在线导入不成功，可选择下载工作流至本地，再利用图 2-21 所示的方法 1 或方法 2 导入。



图 2-22　将 KNIME Hub 工作流拖曳至本地

此外，导入在线共享的工作流组时，不用一个个拖曳工作流至本地。以 KNIME 官方学习材料为例，如图 2-23 所示，该工作流组下子文件夹 Exercises 内有多个工作流，可通过拖动标号 2 工作流至本地，或逐一单击标号 3 中的下载按钮。若想导入在线工作流组，可拖曳该工作流组文件夹图标至本地的 KNIME Explorer 窗格中，如图 2-24 所示。

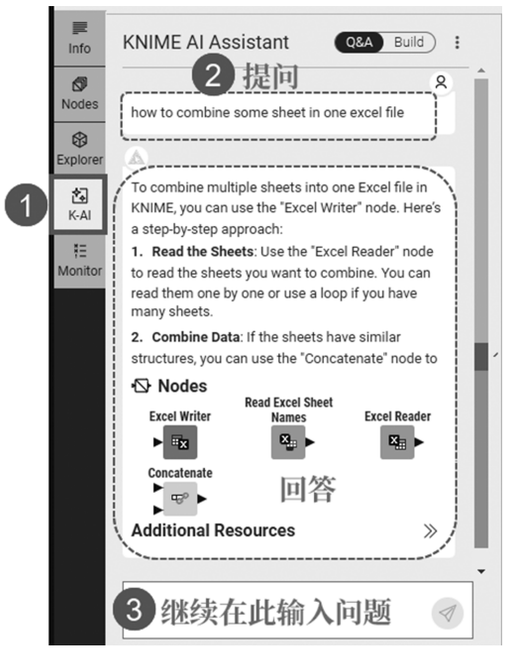


图 2-23　KNIME Hub 在线共享工作流组

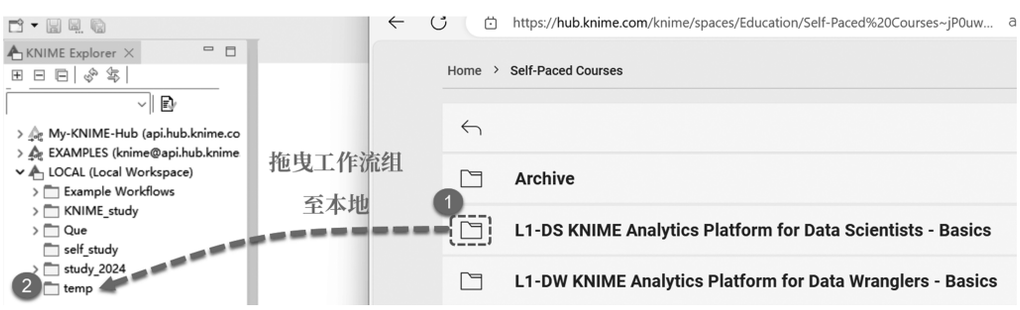


图 2-24　将在线工作流组导入至本地指定位置

2.2.5　工作流其他操作

在经典界面的 KNIME Explorer 窗口内，通过右击选中工作流，可对其进行删除、重命名，复制、剪切、粘贴等操作；还可以拖动一个工作流至其他工作流组中（现代界面下不支持），如图 2-25 所示。相比现代界面，在经典界面中浏览和操作工作流更方便，这也是推荐使用经典界面的原因之一。



图 2-25　工作流复制、移动、粘贴等操作

2.3 工作流注释

注释的主要目的是增强工作流的可读性，方便日后的管理、版本的迭代升级，以及记录关键步骤和方法等，注释主要内容如图 2-26 所示。



图 2-26 工作流注释

KNIME 中，按照注释的层级从小到大排序，可划分为如下四个层级。

（1）单节点注释。在默认情况下，节点信息为“Node 3”之类的序号。若该节点比较关键，涉及关键或复杂的转换，则有必要注释该节点实现了哪些复杂操作，避免日后遗忘。

（2）区域块注释。右击工作流空白区域，选择 New Workflow Annotation，对多个节点实现某个功能进行注释。还有一个类似的操作，通过选中区域内多个节点，右击新建 Metanode，将多个节点转为一个节点，即元节点，如图 2-27 所示。该方式简化了工作流，同样起到类似区块注释的效果。

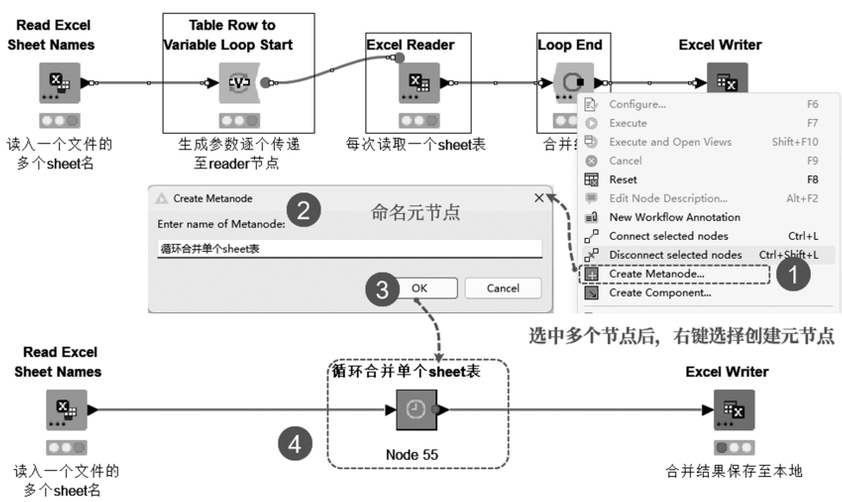


图 2-27 创建元节点（Metanode）

（3）工作流注释。类似区域注释操作，在工作流编辑区节点之外空白处创建注释，主要对整个工作流进行注释和说明，包括工作流的目的、分析思路、关键步骤、注意事项等，以便对整个工作流进行提示和说明。

（4）更多工作流注释。在经典界面下，单击工作流空白区域，点击 Description 右上角钢笔状图标，添加工作流说明、创建时间、作者、URL、标签等，并对整个工作流进行注释。在现代界面下，可点

击软件左侧的 Info 选项中进行工作流注释。

在经典界面下，区域注释和工作流注释的操作便捷性都不如现代界面。现代界面对注释做了很多升级，编辑更加友好，支持项目自动编号，更好地支持文本层级等。一般情况下，创建工作流时可暂不注释，等工作流创建完成能正常运行之后，再对关键节点和关键区域进行必要性注释。注释为用户提供了更好的文档和版本管理。

2.4　数据类型及其转换

2.4.1　数据类型

常见的数据类型有字符串、整型、浮点型等，在 KNIME 数据表中可以看到它们分别对应不同的字母或图标，如图 2-28 所示。此外，在流变量中也需要区分不同的数据类型。数据类型决定了需要使用的节点或方法，如针对数值型节点 Math Formula；若某列是字符串类型，则该节点配置界面中则不显示该列名。数据类型是数据分析的基础，也是 KNIME 软件学习的基础，常用的数据类型如下。

（1）字符串，即 String 类型，所在列以 S 表示。

（2）数值，分为整型 Integer 和浮点型 Double，分别以 I 和 D 表示。

（3）日期时间类型。在查看数据时，其由代表日历的小图标表示。此外，还有代表日期之间差值的 Duration 等类型。

（4）路径，即 Path 类型，主要代表文件的路径，KNIME 早期版本主要由字符串表示。

（5）Collection 类，具体又分为列表 List 和集合 Set，与编程语言中的概念相同。

（6）其他类型，如 JSON、XML、HTML、SVG 图形等类型，不再一一列举。

整型	日期类型	浮点型	字符串	列表	集合
I 编号	日期	D 金额	S 产品	[...] 产品_SplitResultList	[...] 产品_SplitResultSet
1	2023-01-05	396	A, A, C	[A, A, C]	[A, C]
2	2023-01-06	455	C, D, A	[C, D, A]	[C, D, A]
3	2023-01-07	575	A, C, C	[A, C, C]	[A, C]
4	2023-01-08	773	A, B, B	[A, B, B]	[A, B]
5	2023-01-09	501	A, D, E	[A, D, E]	[A, D, E]
6	2023-01-10	337	A, C	[A, C]	[A, C]

图 2-28　不同数据类型的查看方式

2.4.2　数据类型转换

不同的数据类型有不同的操作节点，因此，根据需要进行数据类型之间的转换。常见的数据类型转换如图 2-29 所示。

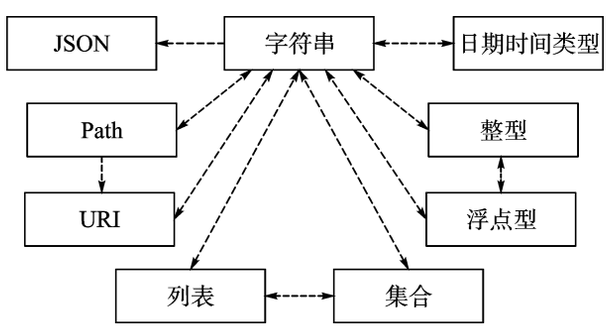


图 2-29　常见不同数据类型的转换

String to Number 和 Number to String 节点可以实现字符串转数值（包括整型和浮点型）和数值转字符串。此外，有些数据类型的转换隐含在节点的函数中，如在 String Manipulation 节点中，函数

toDouble、toInt、string 可实现多个数据类型的转换，如图 2-30 所示。后续章节会根据数据类型的不同，选择相应的节点进行转换和处理。通过图 2-29 中常用数据类型的梳理，结合节点库的节点名，有助于迅速掌握此类节点，4.4 节将深入展开。

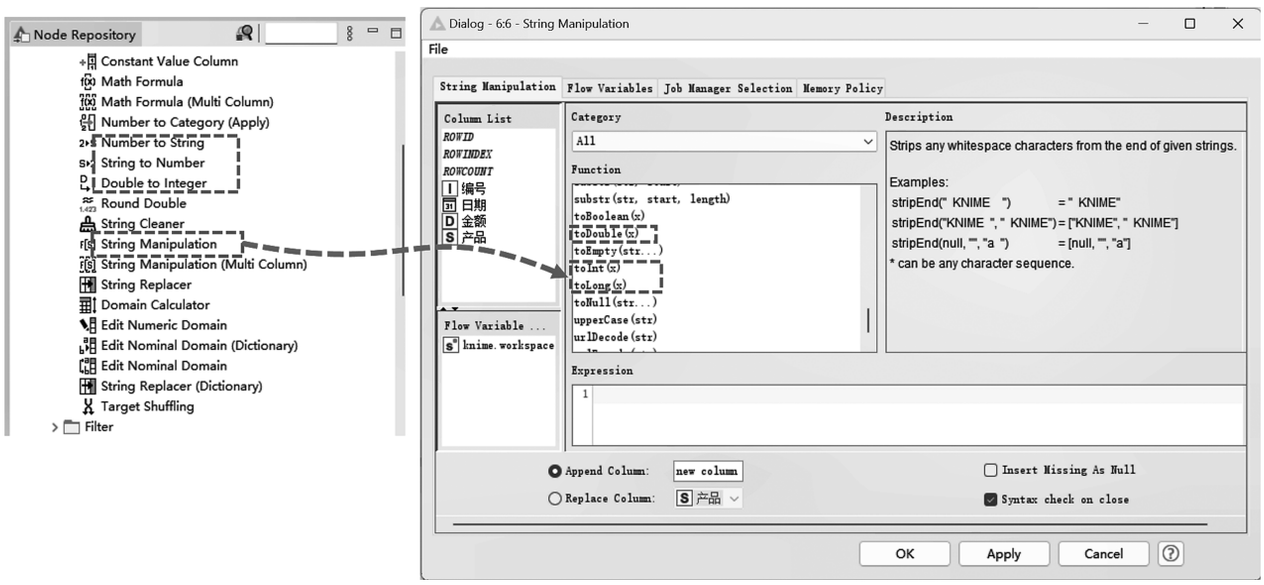


图 2-30　部分数据类型转换节点和函数

2.5　核 心 概 念

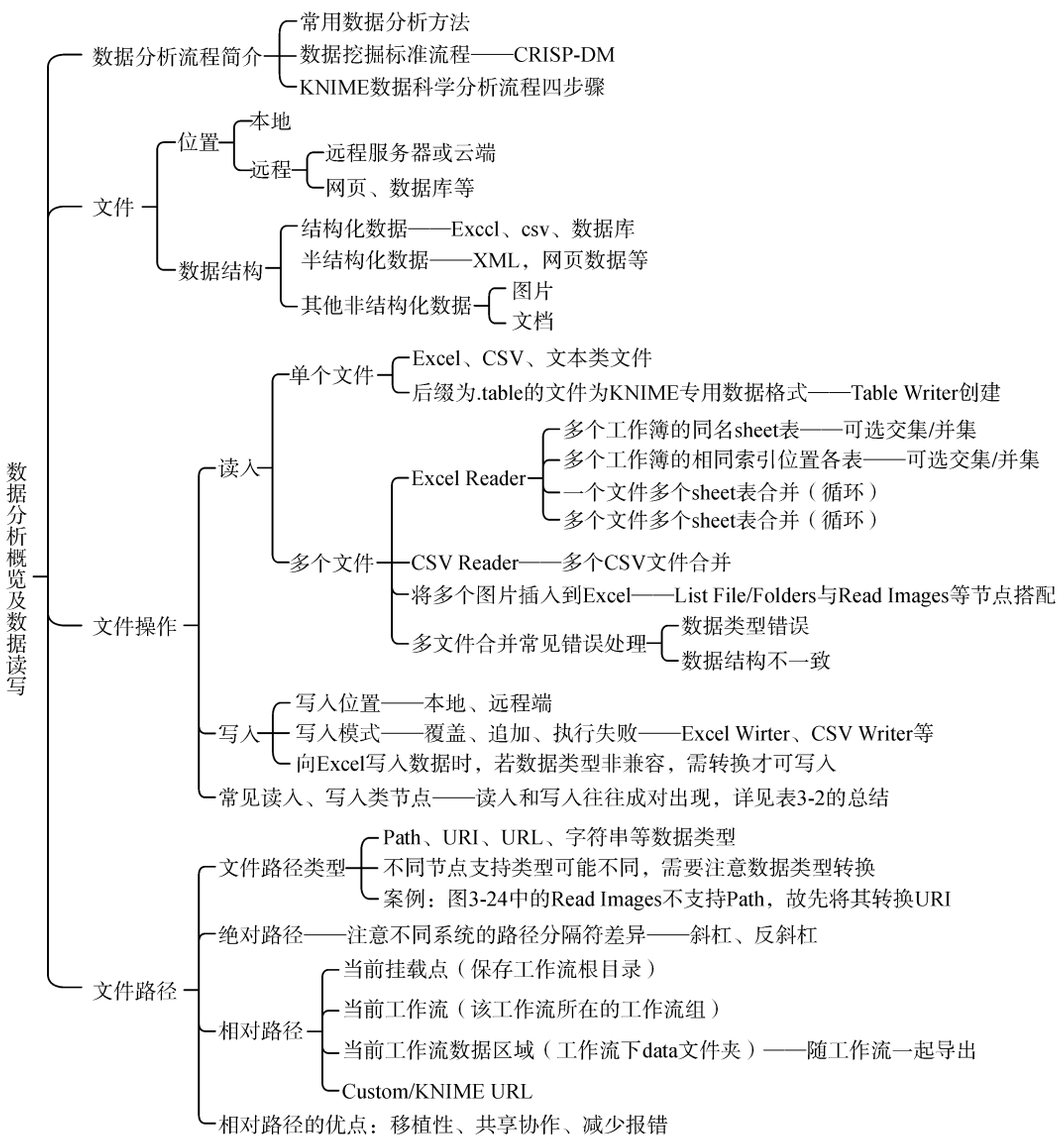
KNIME 中核心概念并不多，主要有以下八个，见表 2-1 所列。

表 2-1　KNIME 核心概念

编号	核心概念	说明
1	Extension	KNIME 扩展，通过安装扩展，可以在节点库中添加更多节点。Extension 类似于编程语言中的 Library（库）、包（Package）、模块（Module）
2	Node	实现某一功能的节点，可类比编程中的函数，一般需要在节点配置选项内设置相关参数。一般情况下，一个扩展包含若干个节点
3	Workflow	工作流，是指在工作流编辑区内将一个个节点有机连接起来，实现特定功能或解决某个问题，是各个节点的有机组合和搭配
4	Workflow Group	工作流组，是存储各个工作流的地方，类似文件夹，通过工作组更好地分门别类地管理工作流。每个工作流组都对应一个本地文件夹
5	Metadata	元数据，是描述数据的数据，即数据表的各列数据类型，包括列序索引、值域范围（Domain）等。每个节点的数据表 Spec 标签页都提供此类信息，或者使用节点 Extract Table Spec 获取数据表中的元数据。更多内容详见本书 6.8 节和 6.9 节
6	Metanode	元节点，将多个连续区域内的节点打包（封装）为一个节点，以简化工作流，使之看起来更整洁，隐藏不必要的细节
7	Component	组件，将多个连续区域内的节点打包为一个节点，除了简化工作流的目的外，组件有多种不同的用途。本书第 13 章至 16 章详细解释组件的不同应用。目的不同，选择的节点也不同，如组件内选择 Configuration 类节点还是 Widget 类节点的目的是不同的。组件属于进阶学习的核心概念之一
8	Flow Variable	工作流变量，简称“流变量”，类似编程语言中的变量概念。在 KNIME 中利用流变量可以实现不同数据间的联动和参数化、自动化，以提高效率。Component 中也不能缺少流变量，它是绝对核心的基础概念。掌握流变量才能理解和学好循环，流变量也是进阶的必备基础，第 9 章会单独展开

第 3 章　数据分析概览及数据读写

本章简要介绍数据分析的常用方法、分析框架和常用数据格式文件的读入类节点，并给出若干个合并多个数据文件的实用场景。当完成数据处理后，使用写入类节点导出相应格式数据。当读写数据文件时，结合 KNIME 对数据文件路径的设置要求，进行详细的图例说明，为后续创建动态路径和文件名打下基础。由于 KNIME 的操作界面基本一致，通过重点节点操作举一反三，可快速掌握各节点功能及重要参数设置。本章主要知识点总结如下。



3.1 数据分析流程简介

对于数据分析的概念，每个人的理解是不同的，每个人面临的业务问题也千差万别，因此很难有统一标准的数据分析流程。对此，业界存在如下不同的标准和方法。

3.1.1 常见数据分析方法

基于应用场景的不同或者使用方法的差异等，结合 KNIME 的功能和对应节点，常见的数据分析方法见表 3-1 所列。

表 3-1 常见的数据分析方法

序号	数据分析方法	功能	对应 KNIME 模块或节点
1	描述性分析	描述数据的基本特征，包括统计指标（如均值、标准差）和可视化（图表展示）	Manipulation 模块、Views 模块、Statistics 节点、Data Explorer 节点
2	诊断性分析	识别数据中的模式和原因，探索变量关系（如相关性、聚类、关联规则）、异常值和缺失值处理	Manipulation 模块、Statistics 子模块、Mining 模块（数据挖掘算法）
3	预测性分析	利用历史数据预测未来趋势，如时间序列 ARIMA、机器学习模型（决策树、随机森林、SVM 等）	Analytics→Mining 子模块、Scikit-learn 扩展、深度学习库 H2O、Keras、TensorFlow 等扩展
4	规范性分析	基于优化理论制定最优策略（如蒙特卡罗模拟、线性规划）	KNIME Optimization Extensions 扩展、Python/R 脚本节点（调用优化库）
5	实验性分析	通过实验设计验证假设，如 A/B 测试、随机对照试验、统计检验如假设检验、方差分析等	Analytics → Statistics 子模块（假设检验、ANOVA 等）
6	文本分析	非结构化文本处理，如情感分析、主题建模、NLP 任务	文本分析扩展、Python Script 节点（调用 NLTK、spaCy 等库）
7	大数据分析	处理海量数据，使用分布式计算框架（如 Hadoop、Spark）	大数据扩展，如 Spark、Hive 等
8	数据库查询	通过 SQL 查询和操作结构化数据库	DB 模块，直连数据库，如 MySQL、PostgreSQL、Oracle、MongoDB 等
9	数据可视化	通过图表（如条形图、折线图、动态图表等）直观展示数据及变化，支持插入报告	Views 模块、Reporting（结合文字和图表生成格式化报告）
10	数据清洗与预处理	清洗数据（处理缺失值、异常值、重复记录）、数据转换（标准化、归一化、独热编码）等	Manipulation 模块，如 Numeric Outliers、Missing Value、Normalizer 节点等

KNIME 中有大量的精细化操作节点，是该软件的一个特色和优势，可以实现数据清洗和预处理。这部分操作可能占据一个分析项目约 80%的时间。KNIME 中的 Manipulation 模块下的众多节点，基本都属于必知必会类节点。

3.1.2 数据挖掘标准流程

CRISP-DM（CRoss-Industry Standard Process for Data Mining，跨行业数据挖掘标准流程），是一种广泛应用于数据挖掘和数据科学项目的标准化流程模型。虽然本书不以数据挖掘为主题，但是该流程具有重要的参考价值，也能为数据分析学习者提供一个指导性框架。

CRISP-DM 主要由六个阶段组成，该流程旨在提供一个结构化的方法来解决数据相关问题。该流程是一个迭代更新的过程，每个阶段都可以根据项目的进展进行调整和回转，如图 3-1 所示。

附录：常用节点名称变化对照表

由于 KNIME 版本在持续更新中，一些重要且常用的节点名称可能被删除、重命名或整合至其他节点中等，读者在学习中可能找不到该同名节点。因此，为方便学习，特将一些重要节点按照节点库的位置，整理成表，以供参考学习。常用节点在 KNIME 4.3 至 5.8 版本中的变化见表附录-1 所列。

表附录-1　常用节点名称变化对照表

序号	所属模块	原名称	新名称	变更类型	版本	备注
1	IO	Zip Files	Compress Files/Folder	重命名	4.3	属于新的文件处理框架
2	IO	Unzip Files	Decompress Files	重命名	4.3	属于新的文件处理框架
3	IO	Create Directory	Create Folder	重命名	4.3	术语统一
4	IO	Create Temp Dir	Create Temp Folder	重命名	4.3	术语统一
5	IO	Delete Files	Delete Files/Folders	重命名	4.3	功能扩展，支持文件夹
6	IO	Copy/Move Files	Transfer Files	整合/重命名	4.3	整合了 Download 和 Upload 节点的功能
7	IO	Download	Transfer Files	整合	4.3	功能并入 Transfer Files
8	IO	Upload	Transfer Files	整合	4.3	功能并入 Transfer Files
9	IO	List Files	List Files/Folders	重命名	4.3	整合了 List Remote Files 功能
10	IO	Excel Reader（XLS）	Excel Reader	重命名	4.3	统一支持 XLS 和 XLSX
11	IO	Excel Writer（XLS）	Excel Writer	整合/重命名	4.3	整合了 Excel Sheet Appender 的功能
12	IO	Excel Sheet Appender	-	删除	4.3	功能已完全整合至新的 Excel Writer 中
13	IO	-	Excel Cell Updater	新增	4.5	在原 Excel 表中修改指定单元格数据
14	IO	-	Example Data Reader	新增	5.3	内置四个示例数据集，方便初学者学习其他节点
15	Manipulation	Cell Replacer	Value Lookup	替代	5.2	被功能更强大的 Value Lookup 节点所替代
16	Manipulation	-	String Cleaner	新增	5.2	新增节点，用于清理字符串
17	Manipulation	Pivoting	Pivot	重命名	5.1	命名规范化，去除“ing”
18	Manipulation	Unpivoting	Unpivot	重命名	5.1	命名规范化，去除“ing”

续表

序号	所属模块	原名称	新名称	变更类型	版本	备注
19	Manipulation	Insert Column Header	Column Renamer（Dictionary）	重命名	5.1	功能描述更准确
20	Manipulation	Extract Column Header	Column Name Extractor	重命名	5.1	命名规范化
21	Manipulation	Row to Column Header	Row to Column Names	重命名	5.1	命名规范化
22	Manipulation	Top k Selector	Top k Row Filter	重命名	5.1	功能描述更准确
23	Manipulation	Transpose	Table Transposer	重命名	5.1	命名规范化
24	Manipulation	Partitioning	Table Partitioner	重命名	5.5	命名规范化
25	Manipulation	Row Sampling	Row Sampler	重命名	5.5	命名规范化
26	Manipulation	Create Table Structure	Table Structure Creator	重命名	5.5	命名规范化
27	Manipulation	Column Rename（Regex）	Column Name Replacer	重命名	5.5	功能描述更准确，增加支持通配符和字面字符替换，不单单支持正则表达式
28	Manipulation	Regex Split	String Splitter（Regex）	重命名	5.3	归类到 String Splitter 系列，功能描述更清晰
29	Manipulation	Column Expressions	Column Expressions（legacy）	标记为 Legacy	5.4	被新版 Expression 节点部分替代，虽被标记 legacy，仍有很多独特功能不可替代
30	Manipulation	Variable Expressions	Variable Expressions（legacy）	标记为 Legacy	5.4	类似 Column Expressions 节点，仅处理流变量
31	Manipulation	–	Expression	新增	5.3	新一代表达式节点，5.5 版本增加对日期和时间的处理。可以引用之前表达式结果
32	Manipulation	–	Variable Expression	新增	5.3	Expression 的流变量版本
33	Manipulation	–	Expression Row Filter	新增	5.3	类似 Expression，基于条件判断真假过滤行，相比 Rule-Based 类行过滤节点可实现更灵活的条件判断
34	Manipulation	Double To Int	Double To Integer	重命名	5.1	完整拼写
35	Manipulation	Round Double	Number Rounder	重命名	5.3	名称更通用
36	Manipulation	Add Empty Rows	Constant Value Row Appender	重命名	5.8	功能不变，界面重新布局。添加空行时，可基于数据类型选择填充常数值
37	Time Series	–	Date Rounder	新增	5.5	将日期圆整至月初/末，季度初/末等，方便聚合操作
38	Time Series	–	Time Rounder	新增	5.5	将时间圆整至最接近的分钟、小时等，方便聚合操作
39	Time Series	Extract Date&Time Fields	Date&Time Part Extractor	重命名	5.3	命名规范化，统一命名为“Extractor”

续表

序号	所属模块	原名称	新名称	变更类型	版本	备注
40	Time Series	Date&Time Shift	Date Shifter / Time Shifter	拆分、重命名	5.5	原节点被拆分为两个独立的节点，分别处理日期和时间
41	Time Series	Modify Date	Date Modifier	重命名	5.5	命名规范化，统一命名为“Modifier”
42	Time Series	Modify Time	Time Modifier	重命名	5.5	
43	Time Series	Modify Time Zone	Time Zone Modifier	重命名	5.5	
44	Time Series	Create Date&Time Range	Date&Time Range Creator	重命名	5.5	命名规范化，统一命名为“Creator”
45	Time Series	Duration to Number	Duration（Time-based）to Number	重命名	5.5	名称更精确
46	Time Series	Extract Duration Fields	Duration Part Extractor	重命名	5.5	命名规范化，统一命名为“Extractor”
47	Time Series	Date&Time Difference	Date&Time Difference	升级功能	5.5	增加选项，明确是列 2 减列 1，或反之
48	Time Series	—	Date&Time Format Manager	新增	5.5	补充日期格式管理。类似 String Format Manager、Number Format Manager
49	Time Series	—	Duration Format Manager	新增	5.5	
50	DB	—	DB Looping	新增	4.6	将选定列的多个唯一属性值替换为 where 条件查询中的占位符
51	DB	—	DB Delete（Filter）	新增	4.6	基于匹配条件移除匹配行，删除原表数据，谨慎操作
52	DB	—	DB Data Spec Extractor	新增	4.6	提取数据库表中的列名、数据类型等元数据
53	DB	—	DB Concatenate	新增	4.6	类似 Concatenate 节点，同 SQL 中的 union 功能
54	DB	DB Column Rename	DB Column Renamer	重命名	5.5	命名规范化
55	DB	DB Column Rename（Regex）	DB Column Name Replacer	重命名	5.5	功能描述更准确，增加支持通配符和字面字符替换
56	DB	DB Table Remover	DB Table Deleter	重命名	5.7	术语统一
57	DB	DB Update	DB Updater	重命名	5.7	术语统一
58	DB	DB Delete（Table）	DB Reference Row Deleter	重命名	5.7	术语统一
59	DB	DB Merge	DB Merger	重命名	5.7	术语统一
60	DB	DB Insert	DB Row Inserter	重命名	5.7	术语统一
61	DB	DB Partitioning	DB Table Partitioner	重命名	5.6	术语统一
62	DB	DB Row Sampling	DB Row Sampler	重命名	5.6	术语统一
63	Workflow Control	Variable to Credentials	Variable To Credentials	重命名	5.1	命名大小写规范化

续表

序号	所属模块	原名称	新名称	变更类型	版本	备注
64	Workflow Control	Variable to Table Column	Variable To Table Column	重命名	5.1	命名大小写规范化
65	Workflow Control	Variable to Table Row	Variable To Table Row	重命名	5.1	命名大小写规范化
66	Workflow Control	Table Row to Variable	Table RowTo Variable	重命名	5.1	命名大小写规范化
67	Workflow Control	Table Column to Variable	Table ColumnTo Variable	重命名	5.1	命名大小写规范化
68	Workflow Control	—	Credentials Properties Extractor	新增	5.3	从凭证流变量中提取非机密信息
69	Workflow Control	—	Variable Filter	新增	5.3	及时过滤不需要的流变量，避免重名冲突
70	Workflow Abstraction	—	Refresh Button Widget	新增	4.4	新增手动重新执行按钮，弥补了部分 Widget 类节点缺少自动重新执行选项
71	Workflow Control	Workflow Service Input	Workflow Input	重命名	5.5	名称简化
72	Workflow Control	Workflow Service Output	Workflow Output	重命名	5.5	名称简化
73	Scripting	Python View	Python View	增加功能	5.7	支持非交互视图插入 Reporting

策划编辑　刘亚莉
责任编辑　刘亚莉
封面设计　墨创文化 M.C CULTURE

KNIME
数据分析
快速入门及进阶



ISBN 978-7-5770-2103-4
9 787577 021034 >
定价: 98.00元