

# KNIME分析平台用户指南



# 目录

介绍.....	1
工作空间.....	2
用户界面.....	3
入口页面.....	5
workflows 编辑器和节点.....	6
连接到KNIME Hub.....	15
切换回KNIME经典用户界面.....	15
空间浏览器.....	17
构建工作流程.....	19
节点仓库.....	20
节点描述.....	22
工作流程描述.....	23
KNIME AI助手.....	23
节点监视器.....	28
帮助.....	30
自定义分析平台.....	30
重置和日志记录.....	30
配置KNIME分析平台.....	31
首选项.....	31
设置knime.ini.....	35
KNIME运行时选项.....	37
KNIME表格.....	44
数据表.....	44
列类型.....	45
排序.....	47
列渲染.....	47
表格存储.....	48
快捷方式.....	50
常规操作.....	50
工作流编辑器模式.....	51
执行.....	51
缩放和平移.....	52
组件和元节点构建.....	53
节点标签.....	54

workflow注释.....	54
快速添加节点.....	56

# 介绍

本指南介绍了KNIME Analytics Platform的基本用法，指导您在平台上的初步步骤，同时提供关于最重要概念的更高级信息，并指示如何配置平台。

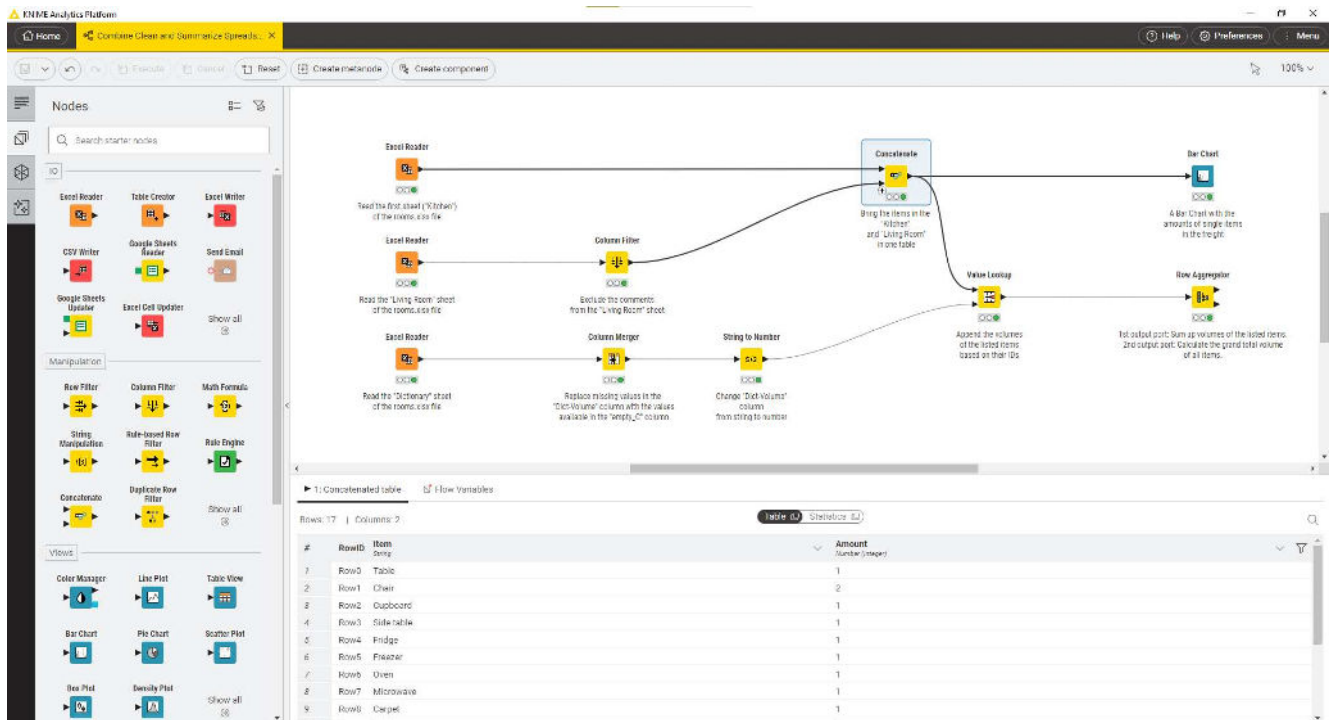


图1. KNIME分析平台

# 工作空间

当您启动KNIME分析平台时，会出现KNIME分析平台启动器窗口并要求您定义KNIME工作区，如图2所示。



KNIME工作区是本地计算机上用于存储KNIME工作流程、节点设置和工作流程生成的数据的文件夹。

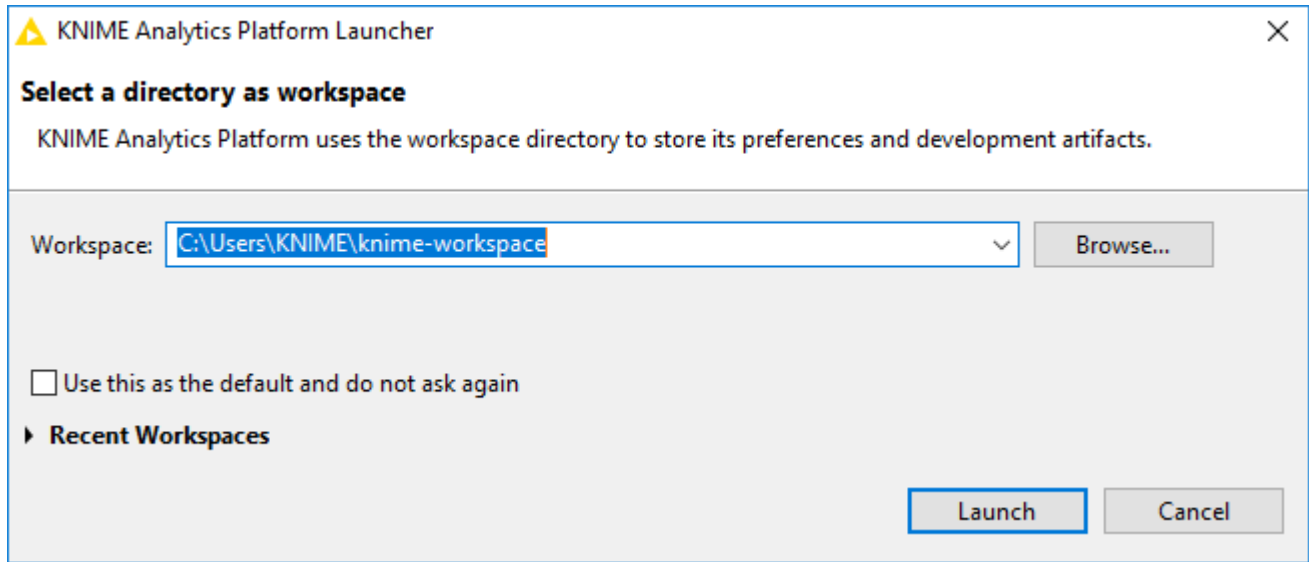


图2. KNIME分析平台启动器

工作区中存储的工作流程、组件和数据可通过侧面板导航中的空间浏览器进行访问。

您可以在稍后的时间在菜单中切换工作区，在用户界面的右上角，选择切换工作区。

# 用户界面

选择当前项目的工作区后，点击启动。 KNIME分析平台用户界面 - KNIME工作台 - 将打开。

从打开的工作流程切换后，KNIME分析平台的活动工作流程将显示出来。 如果在切换视图之前打开了多个工作流程，则只会在KNIME现代界面中显示活动工作流程和所有已加载的工作流程选项卡。切换后，每个工作流程都会显示一个工作流程选项卡。点击第一个选项卡（带有KNIME徽标）后，您将进入入口页面。

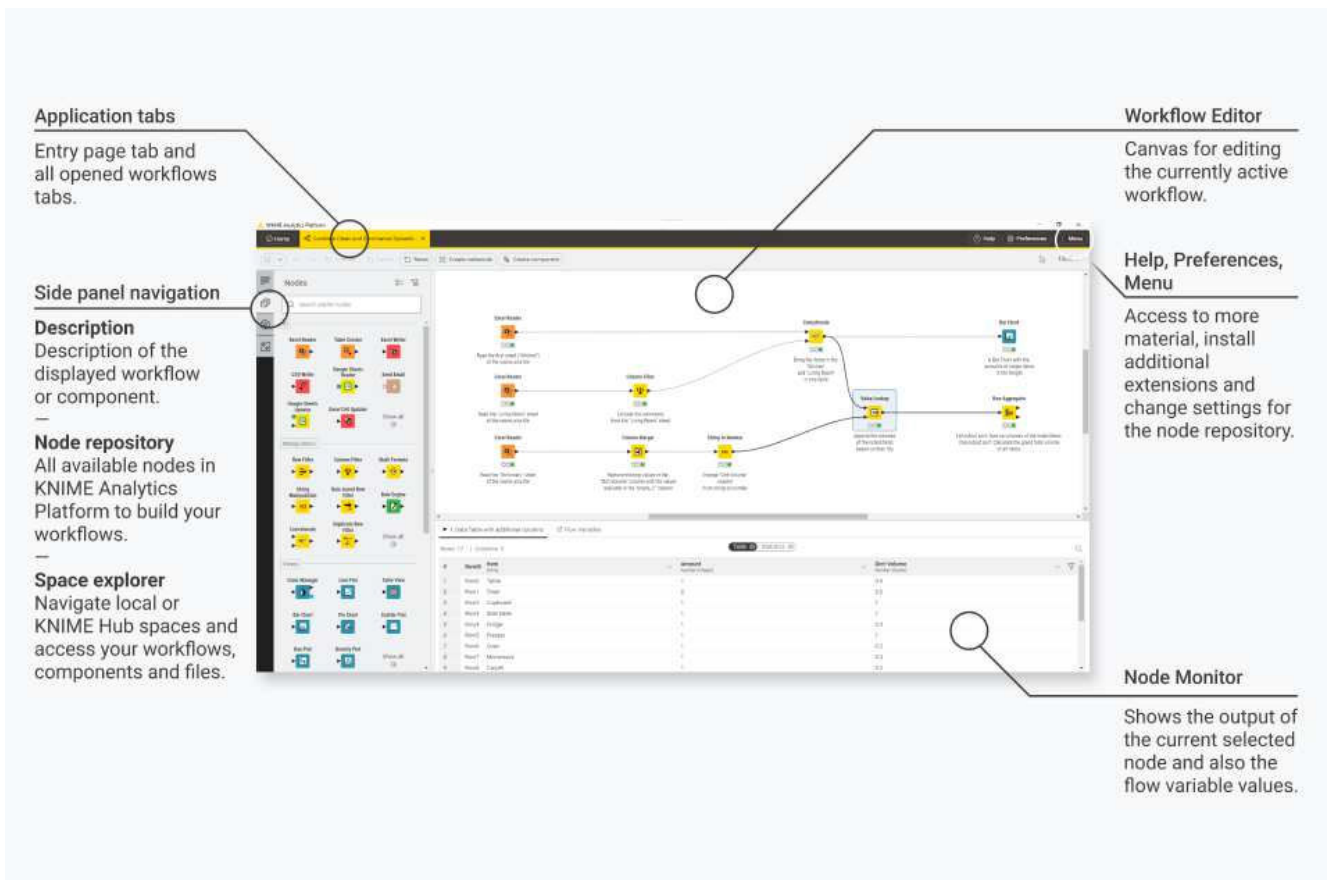


图3. 常规用户界面布局 - 应用程序选项卡、侧面板、工作流编辑器和节点监视器

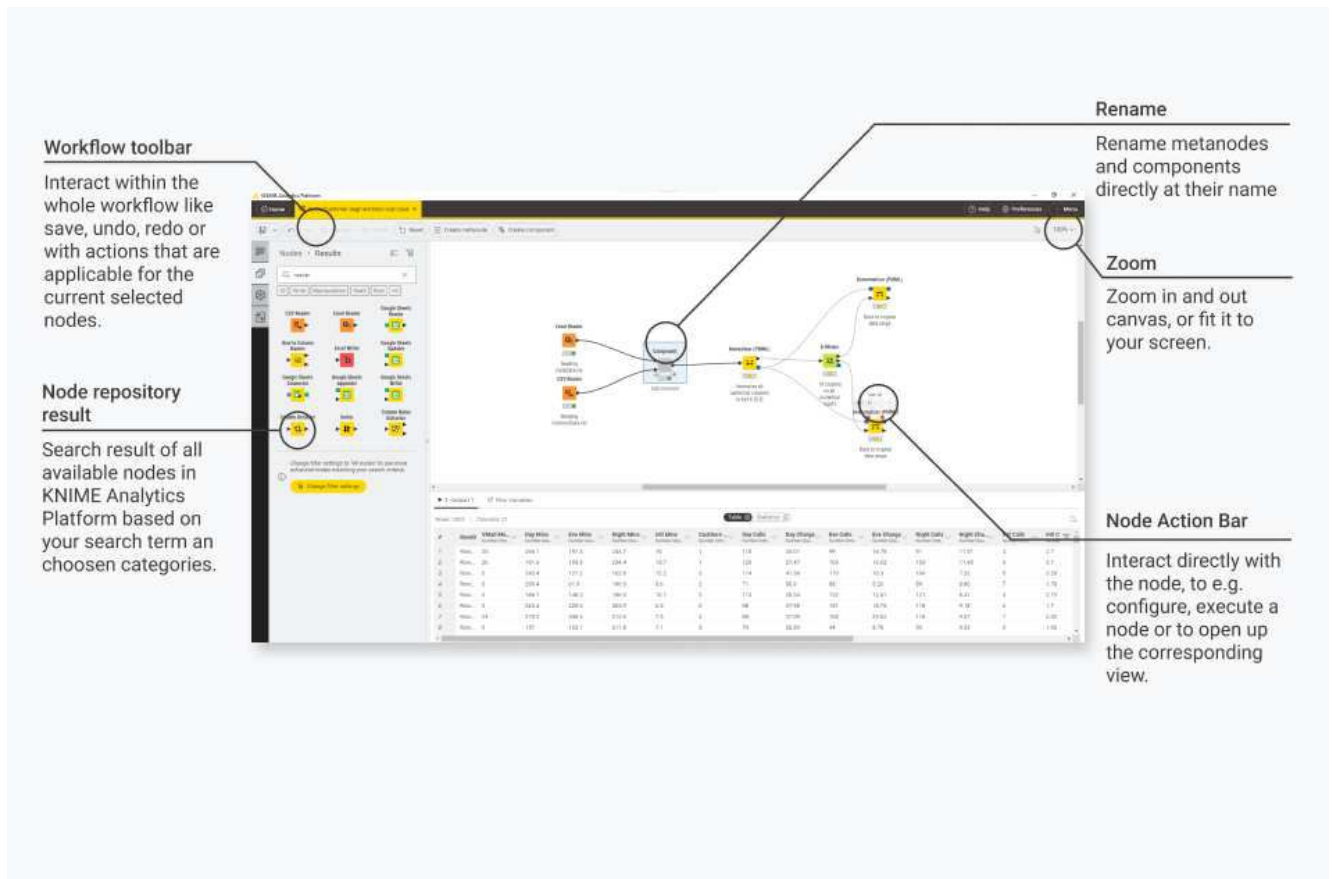


图4. 用户界面元素 - workflow工具栏、节点操作栏、重命名组件和元节点

在接下来的几节中，我们将解释用户界面的这些组件的功能：

- 入口页面
- workflow编辑器和节点
- 连接到KNIME Hub
- 空间浏览器
- 节点仓库
- 节点描述
- workflow描述
- 节点监视器
- 帮助菜单

## 入口页面

单击主页选项卡显示入口页面。

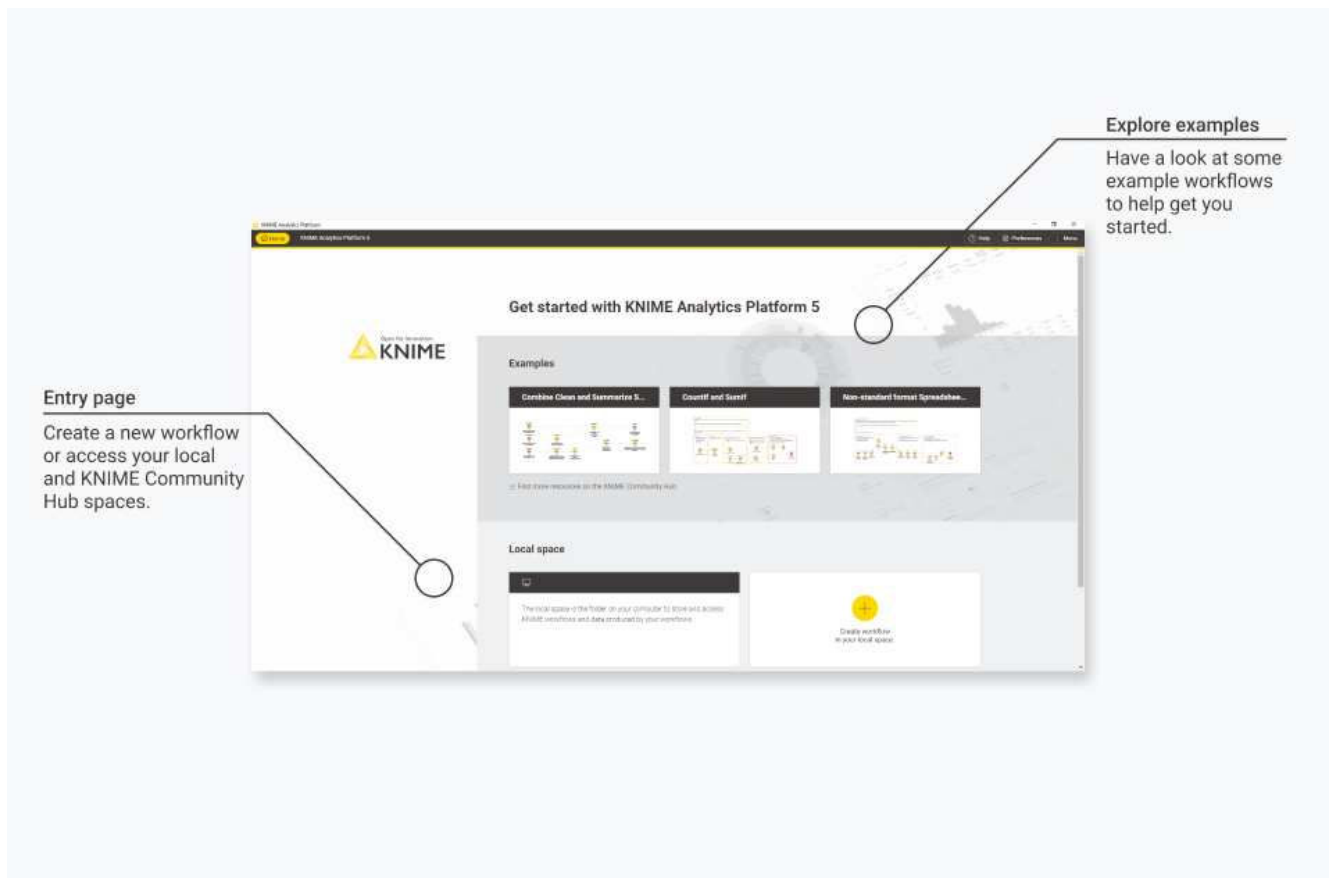


图5. 创建或打开工作流的入口页面

在这里你会找到：

- 三个示例工作流，帮助你入门
- 本地工作空间 - 导航到本地工作空间以找到要处理的工作流
- 点击黄色加号按钮创建新的工作流
- 访问其中一个可用的挂载点。 点击登录，提供您的凭据并开始浏览可用的空间。 默认情况下，只有本地工作空间和连接到个人KNIME社区Hub空间的链接可见。 要添加一个新的挂载点，请按照连接到KNIME Hub部分中的说明操作。



使用KNIME Analytics Platform 5.2版本，您还可以添加KNIME Server挂载点。



## 工作流编辑器和节点

工作流编辑器是组装工作流的地方。工作流由节点表示的单个任务组成。

创建新工作流的一种方法是转到空间浏览器，点击三个点，然后从菜单中选择创建工作流。给工作流命名并点击创建。

在新的空工作流编辑器中，通过从节点库中拖动节点到工作流编辑器中，然后连接、配置和执行它们来创建工作流。

### 节点

在KNIME Analytics Platform中，单个任务由节点表示。节点可以执行各种任务，包括读取/写入文件、转换数据、训练模型、创建可视化等。

#### 节点的相关信息

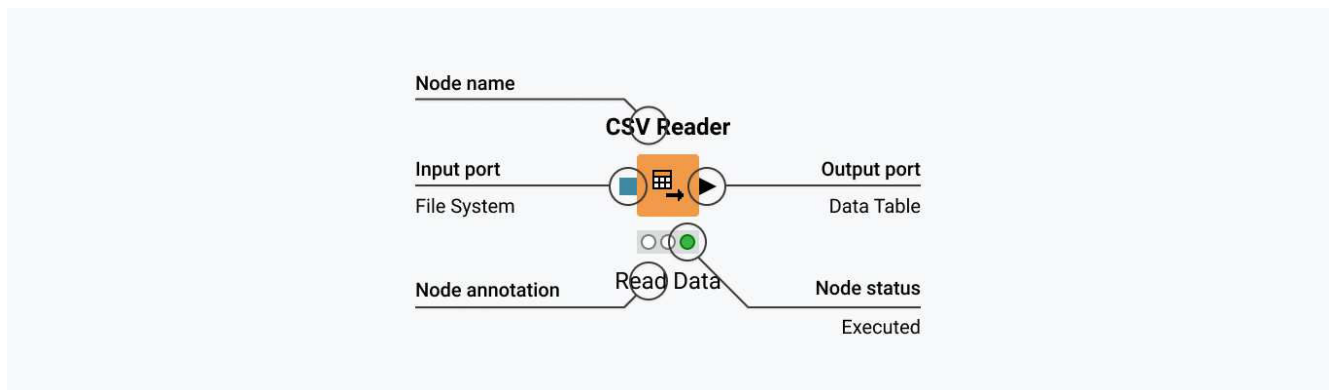


图6。KNIME Analytics Platform中的一个节点

- 每个节点显示为一个带有输入和输出端口以及状态的彩色框，如图6所示。
- 输入端口保存节点处理的数据，输出端口保存操作的结果数据集
- 数据通过连接从一个节点的输出端口传输到另一个节点的输入端口



为简单起见，我们在提到节点的输入和输出端口时称之为数据，但节点也可以有保存模型、数据库查询或其他类型的输入和输出端口，详见节点端口

节点可以处于不同的状态，如图7所示



**Not configured**

The node is waiting for configuration or incoming data.



**Configured**

The node has been configured correctly, and can be executed.



**Executed**

The node has been successfully executed. Results may be viewed and used in downstream nodes.



**Error**

The node has encountered an error during execution.

图7 节点可以存在不同的状态

## 更改节点的状态

可以通过配置、执行或重置来更改节点的状态

所有这些选项可以在以下位置找到：

- 在节点操作栏中 - 单击不同的图标以配置、执行、取消、重置，并在可用时打开视图



图8. 节点的操作栏

- 在节点的上下文菜单中 - 右键单击节点打开上下文菜单

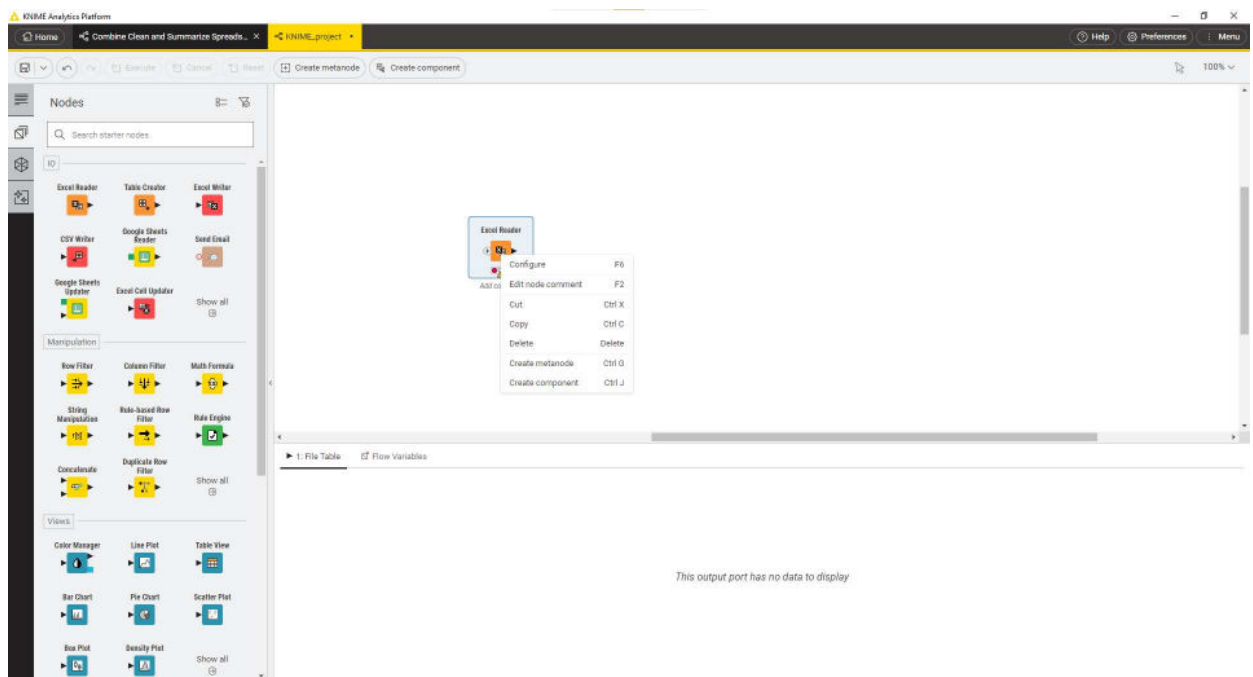


图9. 节点的上下文菜单

## 识别节点状态

每个节点下面的交通灯显示节点的状态。当节点被配置后，交通灯从红色变为黄色，即从“未配置”变为“已配置”。

当新节点首次添加到工作流编辑器时，其状态为“未配置” - 如下所示

通过节点下面的红色交通灯显示。

### 配置节点

可以通过调整配置对话框中的设置来配置节点。

通过以下方式打开节点的配置对话框：

- 双击节点
- 单击节点操作栏中的配置按钮
- 右键单击节点，在上下文菜单中选择配置
- 或者选择节点后按F6键

### 执行节点

有些节点在创建时已经具有“已配置”状态。这些节点在不调整任何默认设置的情况下可执行。

通过以下方式执行节点：

- 单击节点操作栏中的执行按钮
- 右键单击节点并选择执行
- 或者，选择节点并按下F7

如果执行成功，节点状态将变为“已执行”，对应一个绿色的交通灯。如果执行失败，交通灯上将显示一个错误标志，并且必须根据需要调整节点设置和输入。

### 取消节点的执行

要取消节点的执行，请单击节点操作栏中的取消按钮，或右键单击它并选择取消，或选择它并按下F9。

### 重置节点

要重置节点，请单击节点操作栏中的重置按钮，或右键单击它并选择重置，或选择它并按下F8。

。

---



重置节点还会重置工作流中所有后续节点。现在，节点的状态从"已执行"变为"已配置"，节点的输出被清除。

## 节点端口

一个节点可以有多个输入端口和多个输出端口。一组相互连接的节点，使用左侧的输入端口和右侧的输出端口，构成一个工作流。输入端口从前置节点的输出端口消耗数据，输出端口向后续节点提供数据。

除了数据表，输入和输出端口还可以提供其他类型的输入和输出。对于每种类型，输入和输出端口的配对看起来都不同，如图10所示。

输出端口只能连接到相同类型的输入端口 - 数据到数据，模型到模型等等。

某些输入端口可以为空，例如图10中决策树视图节点的数据输入端口。这意味着输入是可选的，节点可以在没有输入的情况下执行。必填输入由填充的输入端口表示，必须提供这些输入才能执行节点。

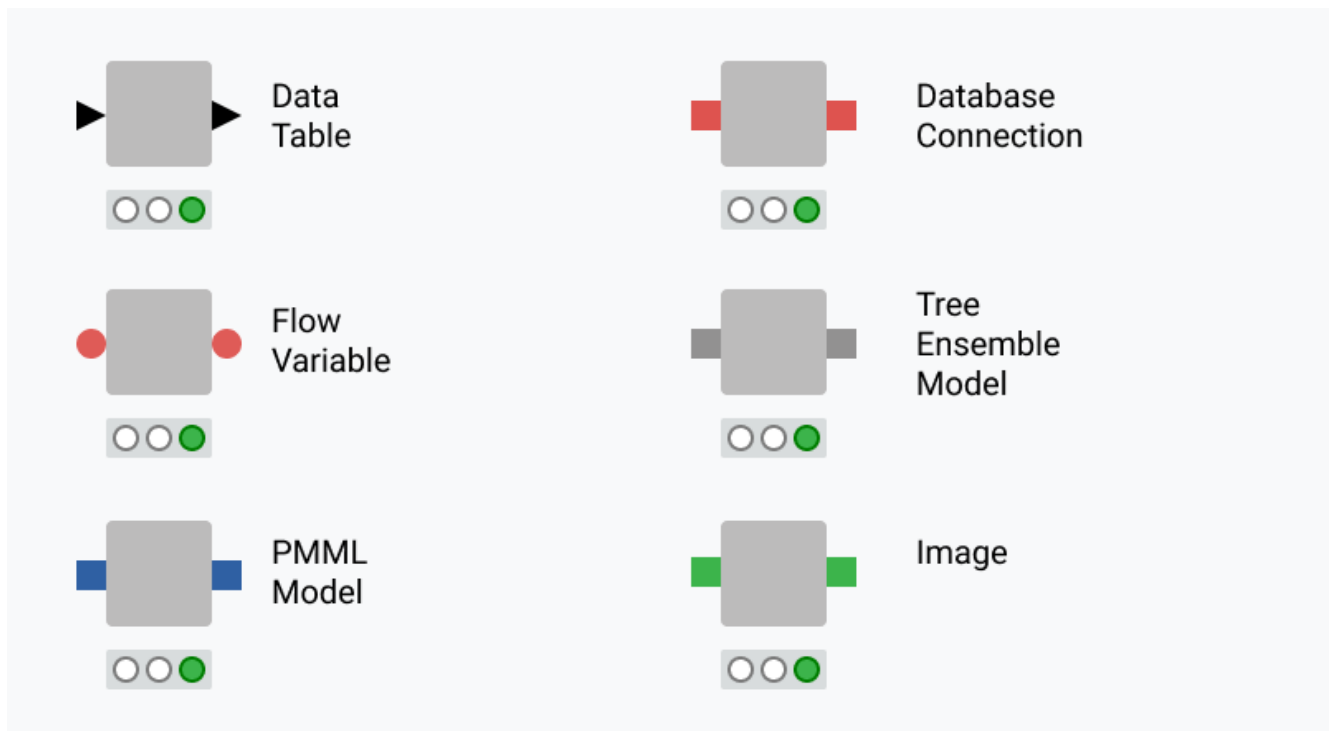


图10. 常见端口类型

工具提示提供了输入和输出端口的简短解释。如果节点被执行，

输出数据的维度将显示在其数据输出端口中。节点描述中有关输入和输出端口的更详细解释。

## 将节点添加到画布

目前有三种方法可以将节点添加到画布以构建工作流程：

1. 从节点库中拖放一个节点,
2. 双击节点库中的一个节点，或者
3. 将连接拖放到 workflow 画布的空白区域以显示快速节点添加面板。此面板内显示最多12个推荐节点。同时你可以在面板中搜索所有兼容的节点。点击所需节点以添加。

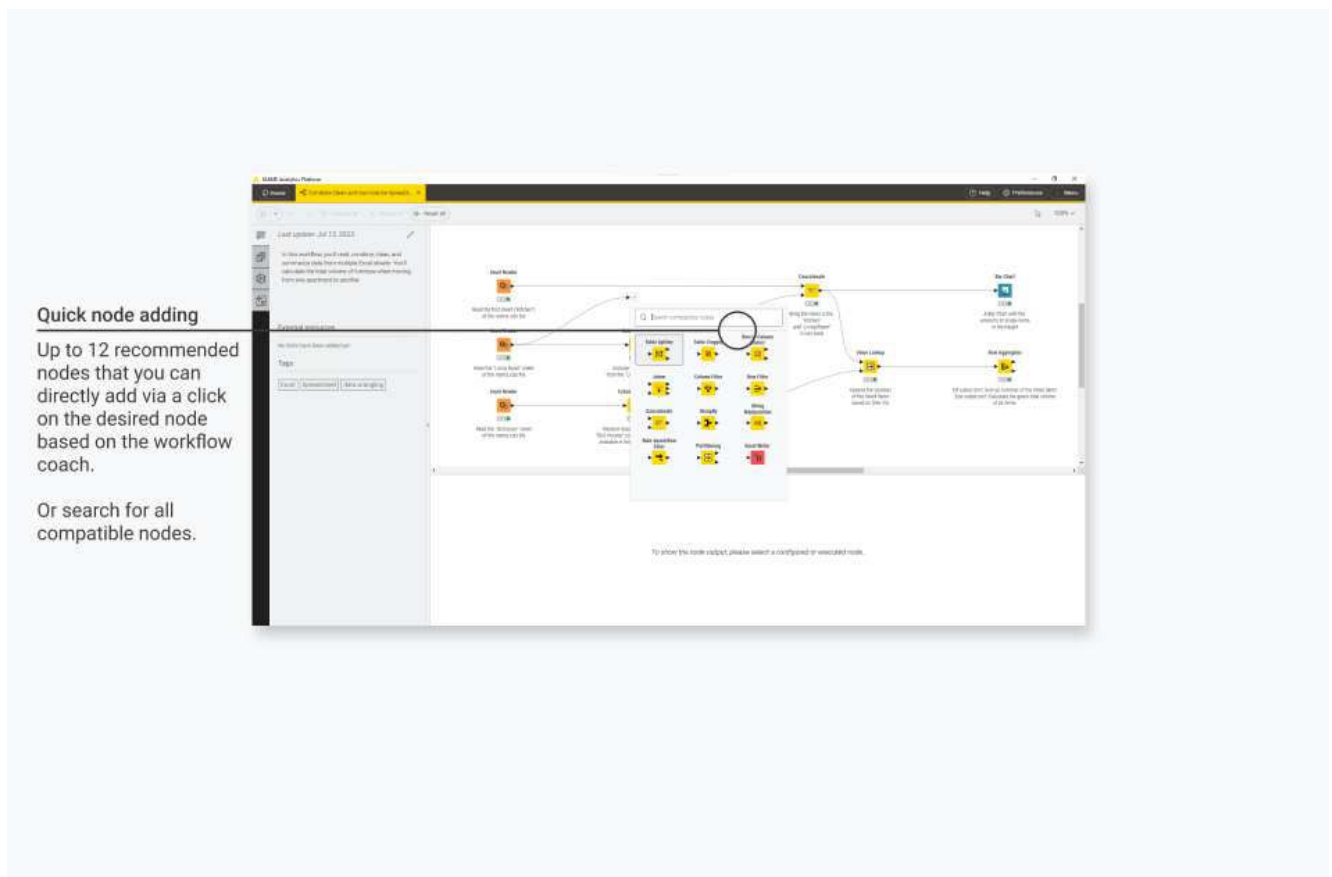


图11. 使用推荐节点的快速节点添加

要使用快速节点添加功能，您需要允许我们接收匿名使用数据。在启动KNIME Analytics Platform时或在切换到新的工作区后，通过在“帮助改进KNIME”对话框中选择是来实现。

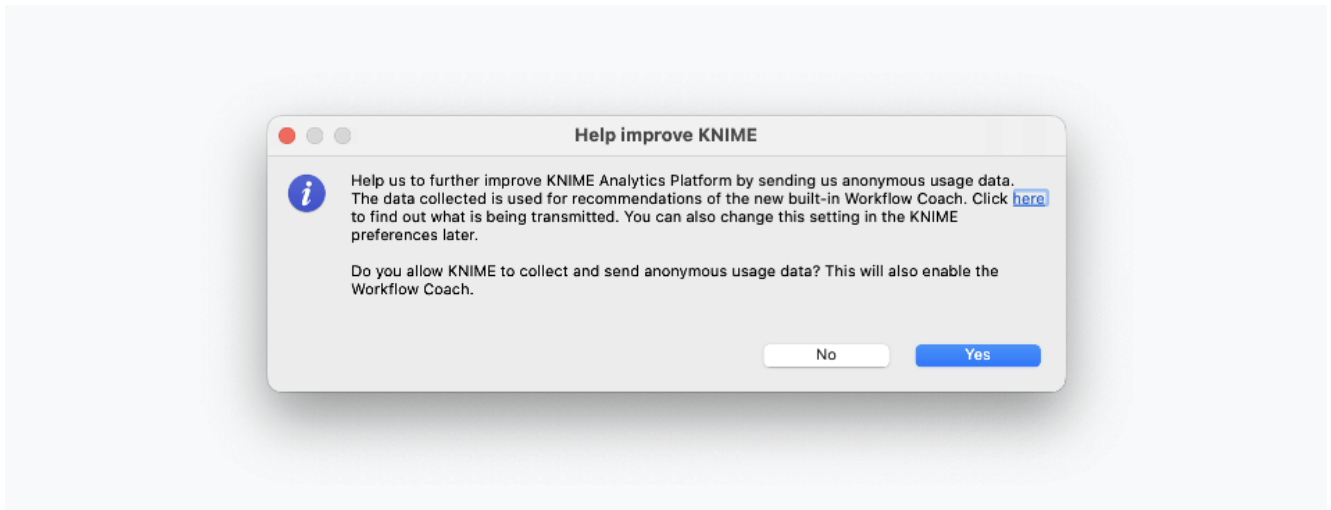


图12. “帮助改进KNIME”对话框

您还可以通过快速节点添加面板中显示的打开首选项按钮来激活它。

点击[此处](#)查看正在传输的内容。如果您不再希望这样做，可以随时在KNIME工作流教练首选项中将其停用。

要打开首选项，请按照以下步骤进行：

1. 在用户界面的右上角点击首选项
2. 转到 **KNIME** → 工作流教练
3. 取消激活社区节点推荐设置

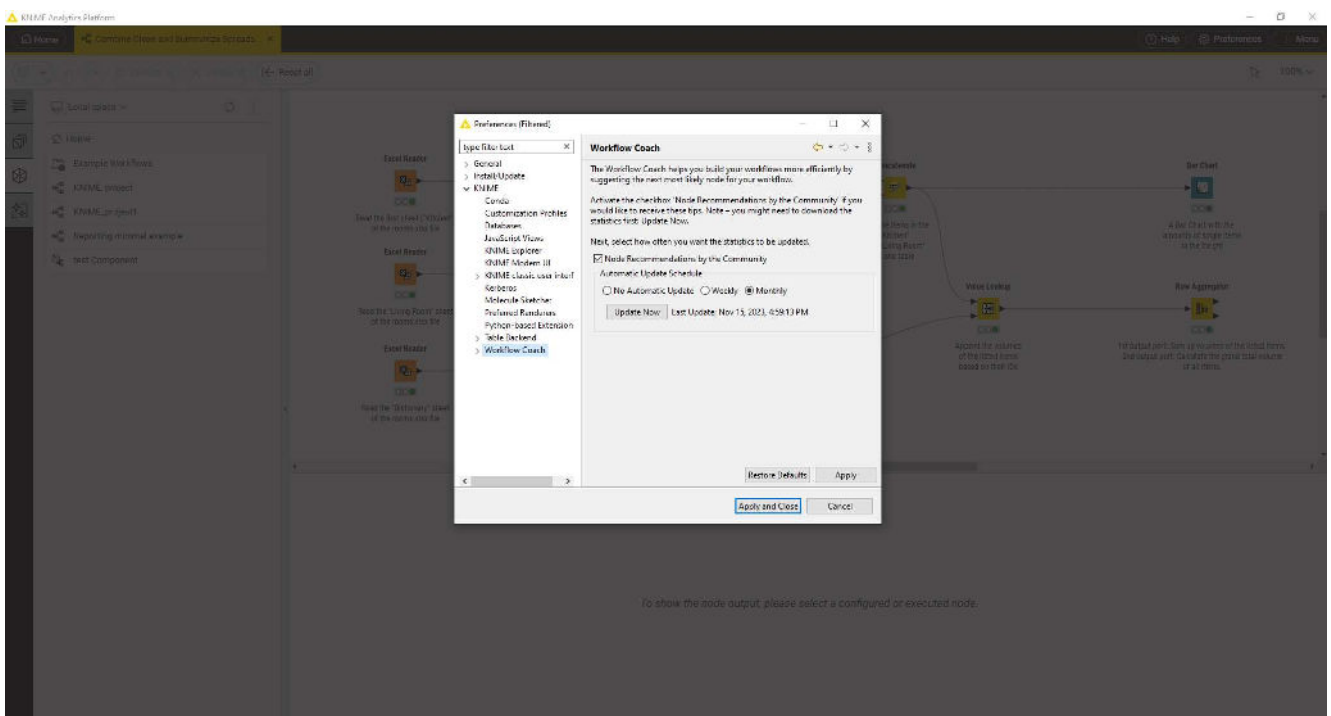


图13. 工作流教练首选项

## 如何选择、移动、复制和替换工作流中的节点

可以通过拖放将节点移动到工作流编辑器中。要在工作流之间复制节点，请选择所选节点，右键单击选择，然后在菜单中选择复制。在目标工作流中，右键单击工作流编辑器，然后在菜单中选择粘贴。

要在工作流编辑器中选择一个节点，请单击一次，它将被一个边框包围。要选择多个节点，请使用鼠标在节点上绘制一个矩形。

通过将新节点拖放到现有节点上来替换节点。现在，现有节点将被一个带有箭头和方框的彩色框覆盖，如图14所示。释放鼠标将替换节点。

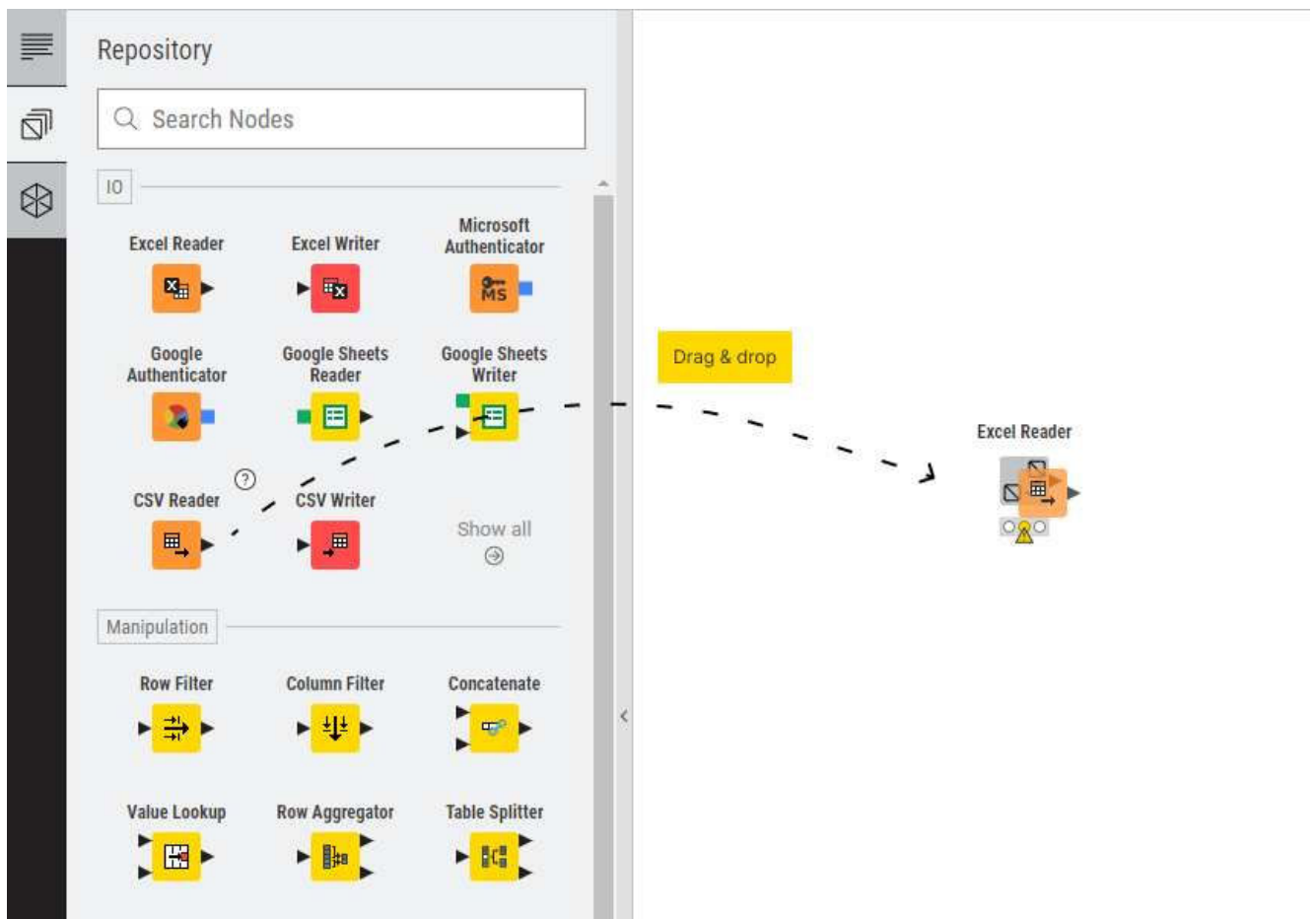


图14。在工作流中替换节点

## 注释和批注

在工作流编辑器中，您有两个选项来记录工作流：

- 节点标签 - 通过双击节点下方的文本字段并编辑文本，向单个节点添加注释



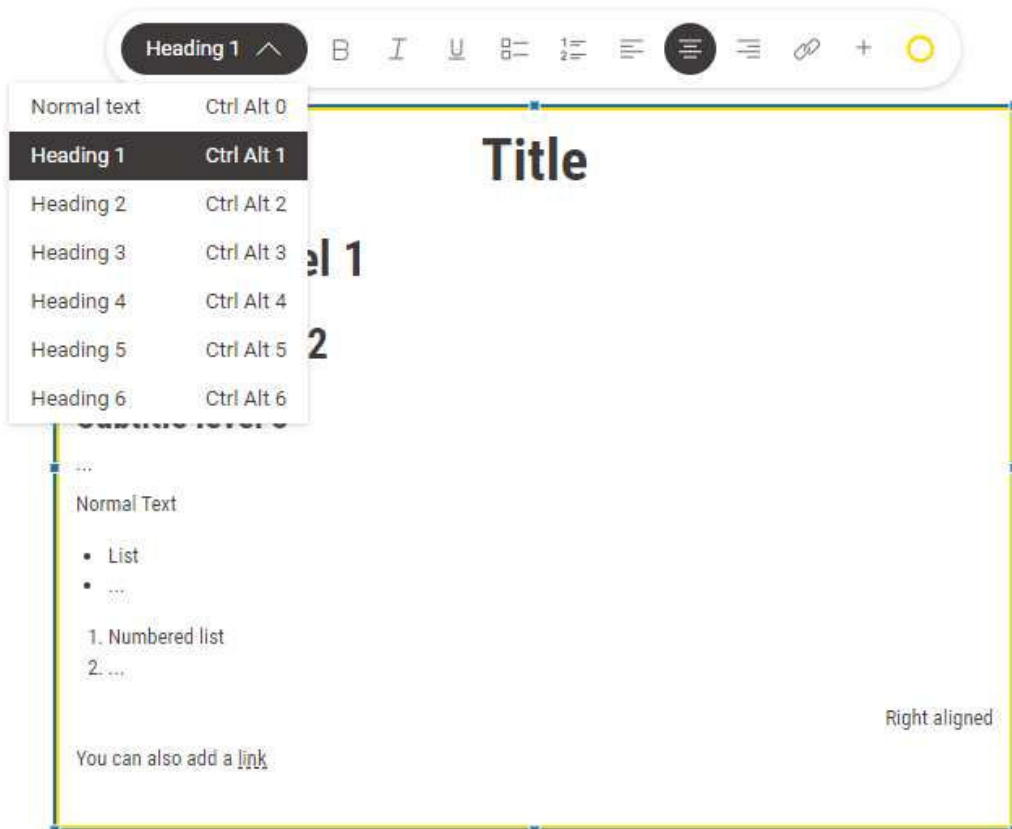


图15. 编写节点注释

- 工作流注释 - 在工作流编辑器中右键单击工作流，然后在菜单中选择新建工作流注释。现在，在工作流编辑器中将出现一个文本框。

i

要添加新的注释，您还可以通过点击用户界面右上角的图标并选择注释模式，或者按下 **T** 键进入注释模式。



双击工作流注释以添加文本、格式化文本和更改注释轮廓的颜色。要更改格式，您可以使用注释栏或使用以下语法：

- 要创建标题，在单词或短语前面添加井号（#），后跟一个空格。您使用的井号数量应与标题级别（<h1>到 <h6>）相对应。
- 要创建项目符号列表，请添加星号（\*）后跟一个空格。
- 要创建一个编号列表，请添加一个数字，后面跟着一个点（ 1.），然后是一个空格。
- 要使文本加粗、斜体或下划线，请选择文本，然后按下CTRL+b、CTRL+i、CTRL+u。

最后，您可以单击注释外部，然后再次单击注释以在画布上移动它或更改其尺寸。

## 连接到KNIME Hub

默认情况下，您可以从主页选项卡连接到KNIME社区中心的帐户。

您可以通过单击首选项，在用户界面的右上角，添加新的KNIME Hub实例挂载点。

转到KNIME资源管理器部分，然后单击新建...。在打开的窗口中，选择KNIME Hub并添加您的Hub URL。然后单击应用。

现在，新的挂载点将显示在主页选项卡中。

登录并选择要操作的空间。空间的内容和相关的操作可在空间资源管理器中看到。

## 切换回KNIME经典用户界面

您可以在用户界面右上角的菜单中选择切换到经典用户界面来切换回经典KNIME Analytics Platform用户界面。

你可以随时按下经典用户界面右上角的打开KNIME现代用户界面按钮来切换回KNIME现代用户界面。

---

非常，非常，非常重要的免责声明

**i**

工作流元素（如连接器或注释）以一种新的方式进行可视化，可能与当前的KNIME Analytics Platform不完全相同。因此，更改不会完全相同。

## 空间浏览器

空间资源管理器是您可以在其中管理工作流、文件夹、组件和文件的地方，可以是本地空间，也可以是KNIME Hub实例上的远程空间。一个空间可以是：

- 您在KNIME Analytics Platform启动时选择的本地工作空间
- KNIME社区Hub上的一个用户空间
- KNIME商业Hub上的一个团队空间



请注意，如果您想使用KNIME服务器工作，您需要切换到经典用户界面。

您可以通过以下方式切换到其他空间：

- 转到主页选项卡，并选择一个可用的空间
- 在空间浏览器的顶部，您可以登录到任何Hub或Server挂载点并选择一个空间。 当在KNIME Hub上时，您将看到按所有者分组的空间。

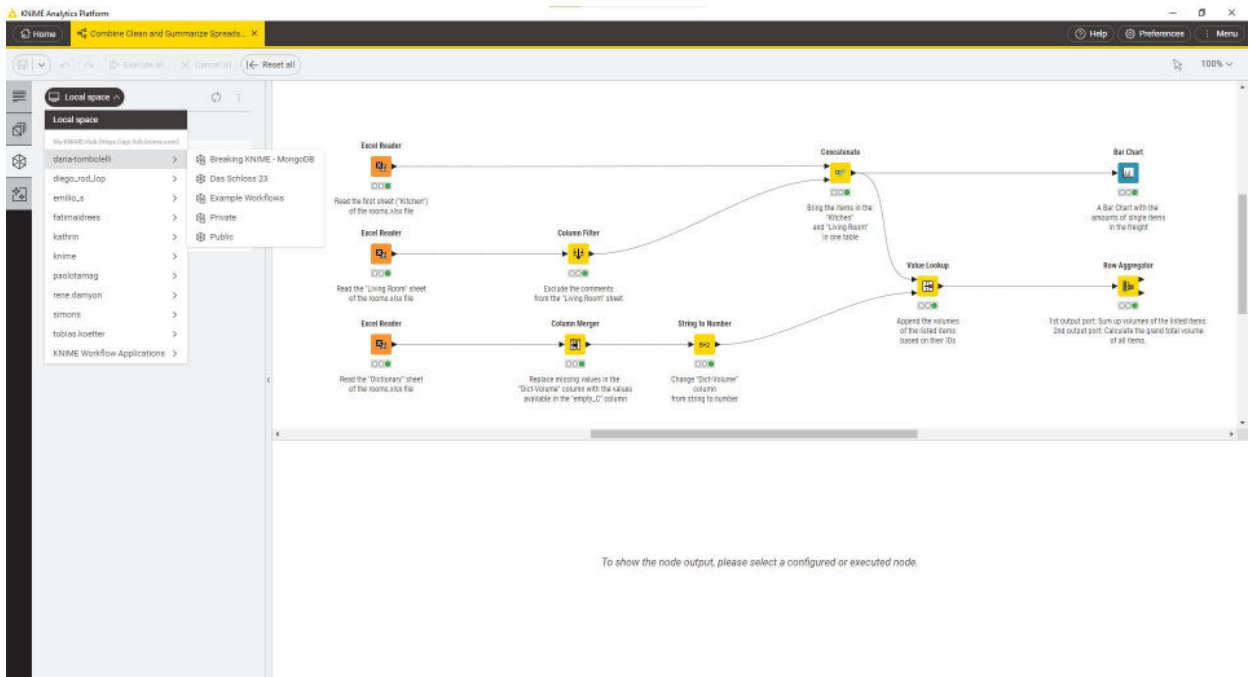


图16。 选择要浏览的空间

在空间浏览器中，您可以看到：

- 工作流程
- 文件夹
- 数据文件

- 组件
- 元节点

双击一个新的空白工作流程，在工作流程画布中打开它，并从节点库开始向画布添加节点。



有关组件和元节点的概述，请参阅KNIME [组件指南](#)。

在此处，您可以点击三个点以在当前空间中选择以下操作之一：

- 创建新文件夹或新工作流程
- 导入工作流程
- 添加文件

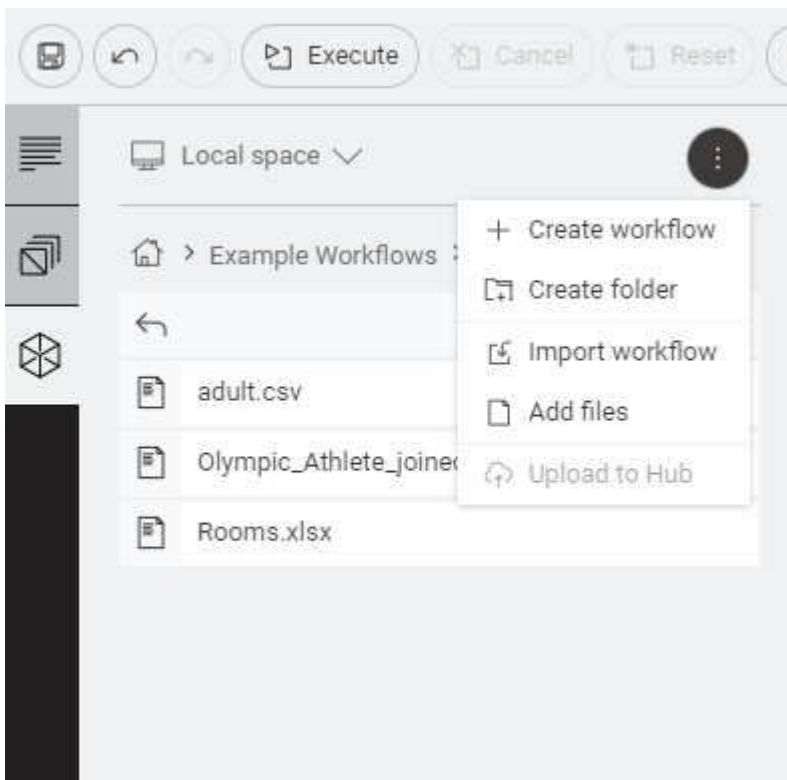


图17. 空间上下文菜单

您还可以将文件拖放到画布上。KNIME将自动创建适当的文件读取节点并预配置它。最后，您可以将组件拖放到画布上以在当前工作流程中使用该组件。

从当前空间中选择一个项目，右键单击它以访问项目上下文菜单。

从这里，您可以重命名、删除、导出（仅适用于工作流程）、上传（如果您已连接到可用的Hub挂载点之一）或连接。

## 构建工作流程

创建新工作流程时，画布将为空。

要构建工作流程，您需要从节点存储库中拖放节点并连接它们。或者，您可以拖动节点的输出端口以显示工作流程教练，它将为您建议兼容的节点并直接连接它们。

一旦将两个节点添加到工作流程编辑器中，您可以通过单击第一个节点的输出端口并将鼠标释放到第二个节点的输入端口来连接它们。

现在，节点已连接。对于某些节点，您可能具有添加特定端口的能力。当将鼠标悬停在这些节点上时，会出现一个 + 符号。单击它以添加一个端口。

如果节点支持不同类型的这些动态端口，将会出现一个列表供您滚动选择要添加的端口类型。

您还可以在工作流中的两个节点之间添加一个节点。要这样做，请从节点存储库中拖动节点，并将其释放到工作流中的相应位置。

---

## 节点仓库

当前安装的节点可在节点存储库中使用。您可以通过拖放的方式将节点从节点存储库添加到工作流编辑器中，如“构建工作流”部分所述。

通过在节点存储库顶部的搜索字段中键入搜索词来搜索节点，如图18所示。

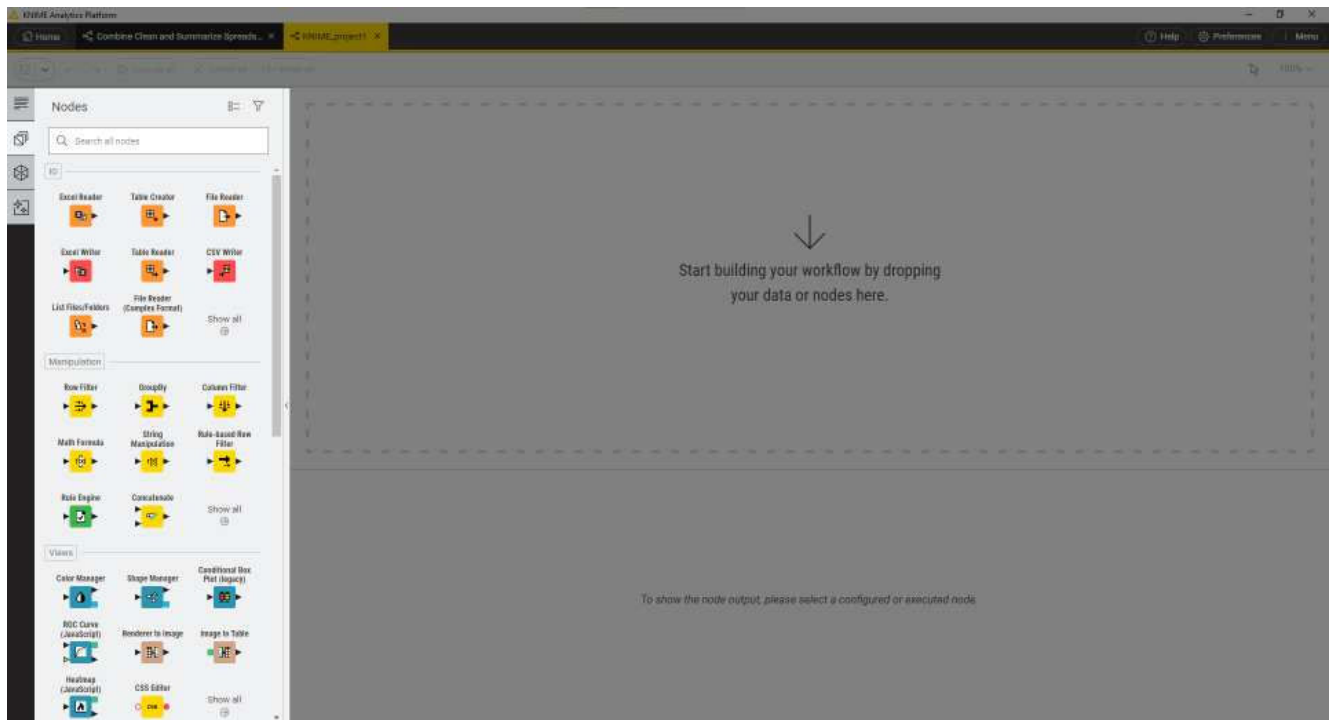


图18。具有两种搜索模式的节点存储库

默认情况下，将显示一组特定的节点，以帮助您开始使用KNIME Analytics Platform。您可以通过更改节点存储库的过滤器设置来扩展搜索结果。单击节点存储库中的图标，进入KNIME Modern UI首选项页面，并更改节点存储库搜索结果中包含的节点的默认设置。

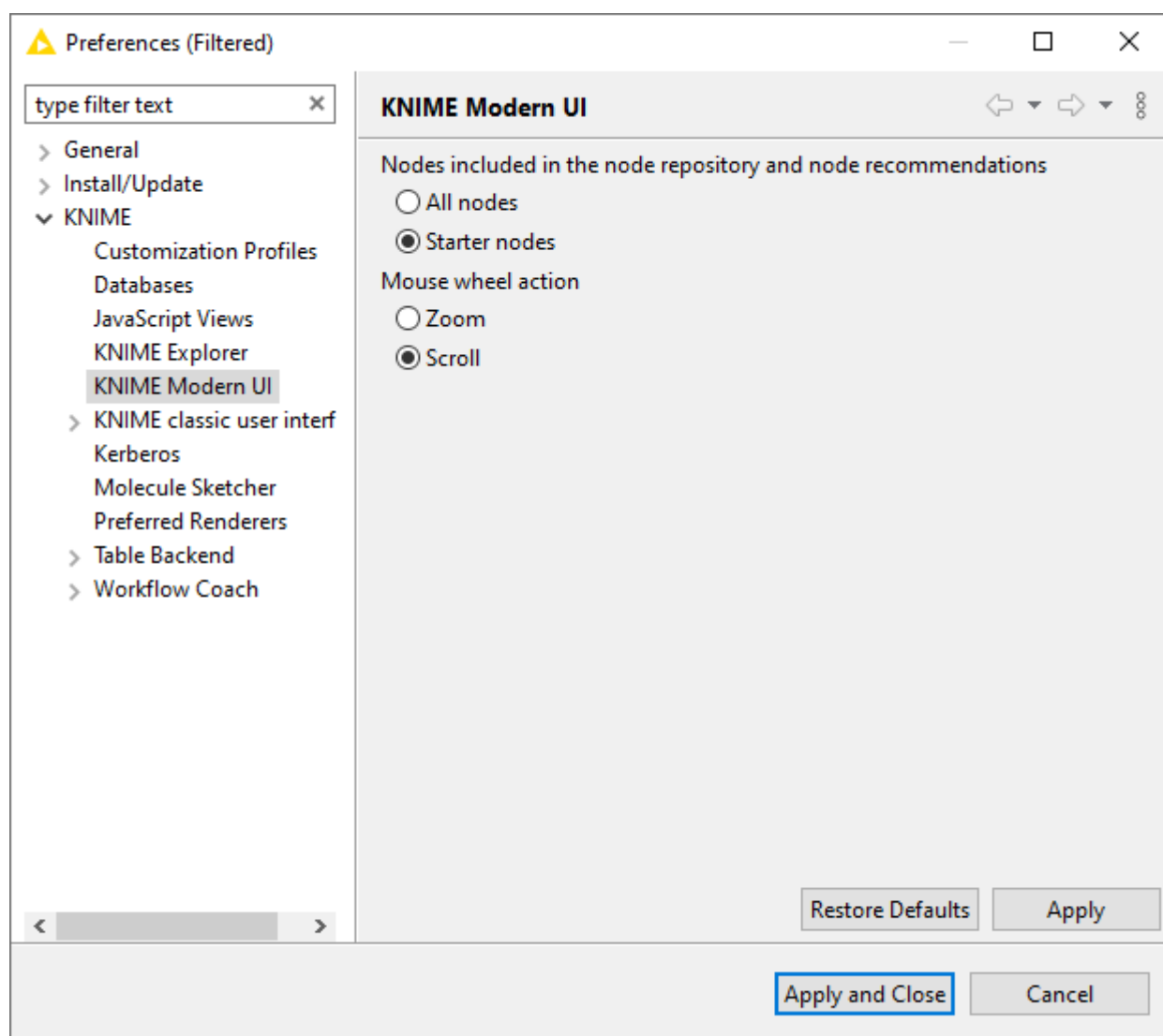


图19. 更改节点搜索结果的默认设置

您可以将节点存储库显示为图标或列表视图。要在这两种状态之间切换，请单击（图标视图）或（列表视图）。



## 节点描述

您可以通过以下方式访问节点描述，了解节点功能、节点配置以及节点可用的不同端口：

- 选择您在画布中添加的节点，转到侧面板导航并选择第一个选项
- 将鼠标悬停在节点存储库中的节点上，然后单击出现的信息图标。这将打开节点描述面板。

## workflows 描述

KNIME Analytics Platform左侧的描述面板提供了当前活动工作流或选定组件的描述。

单击铅笔图标以更改工作流描述，添加外部资源链接和标签。

## KNIME AI助手

KNIME具有一个AI助手，旨在高效回答您对KNIME平台的查询，并协助构建定制工作流，简化数据分析任务。

## 安装

要安装AI助手，请首先在KNIME界面中找到并打开AI助手侧边栏，如图20所示。然后，点击“安装AI助手”按钮，并按照安装菜单中的提示仔细完成设置过程。

如果您找不到AI助手侧边栏，则可能是您的管理员停用了AI助手。

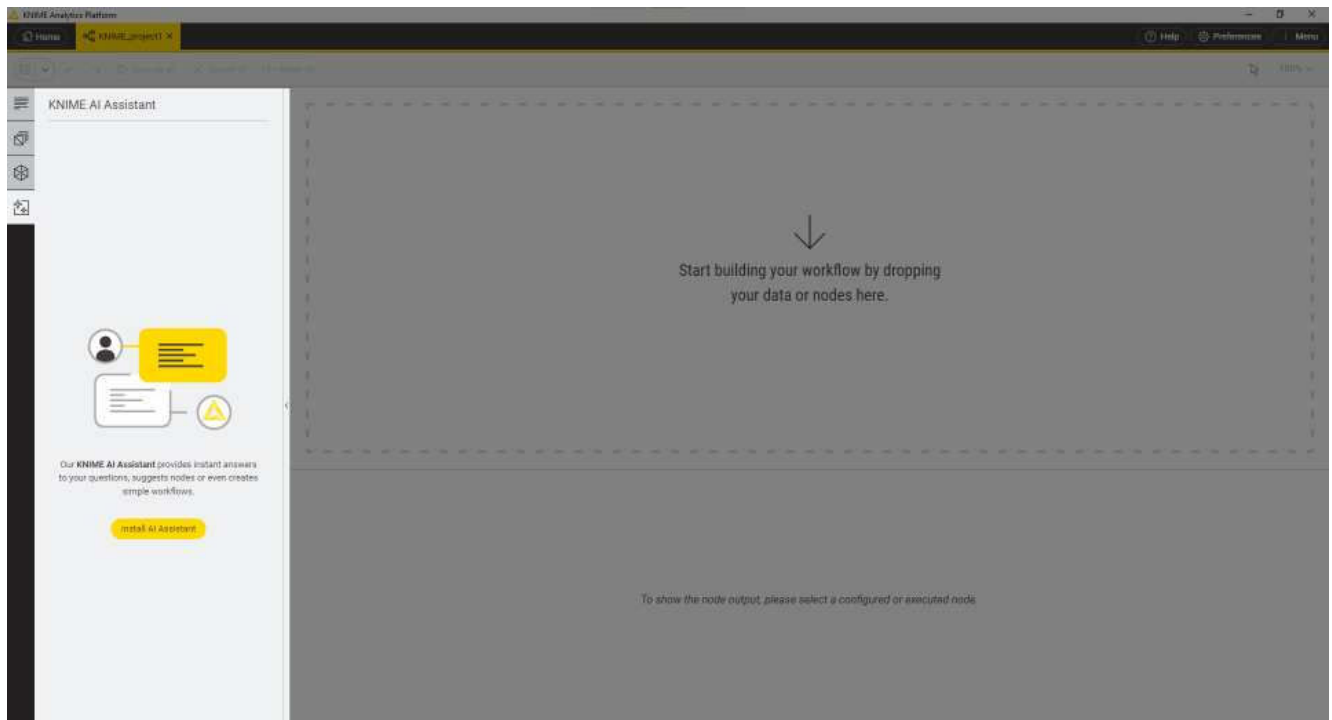


图20. AI助手侧边栏

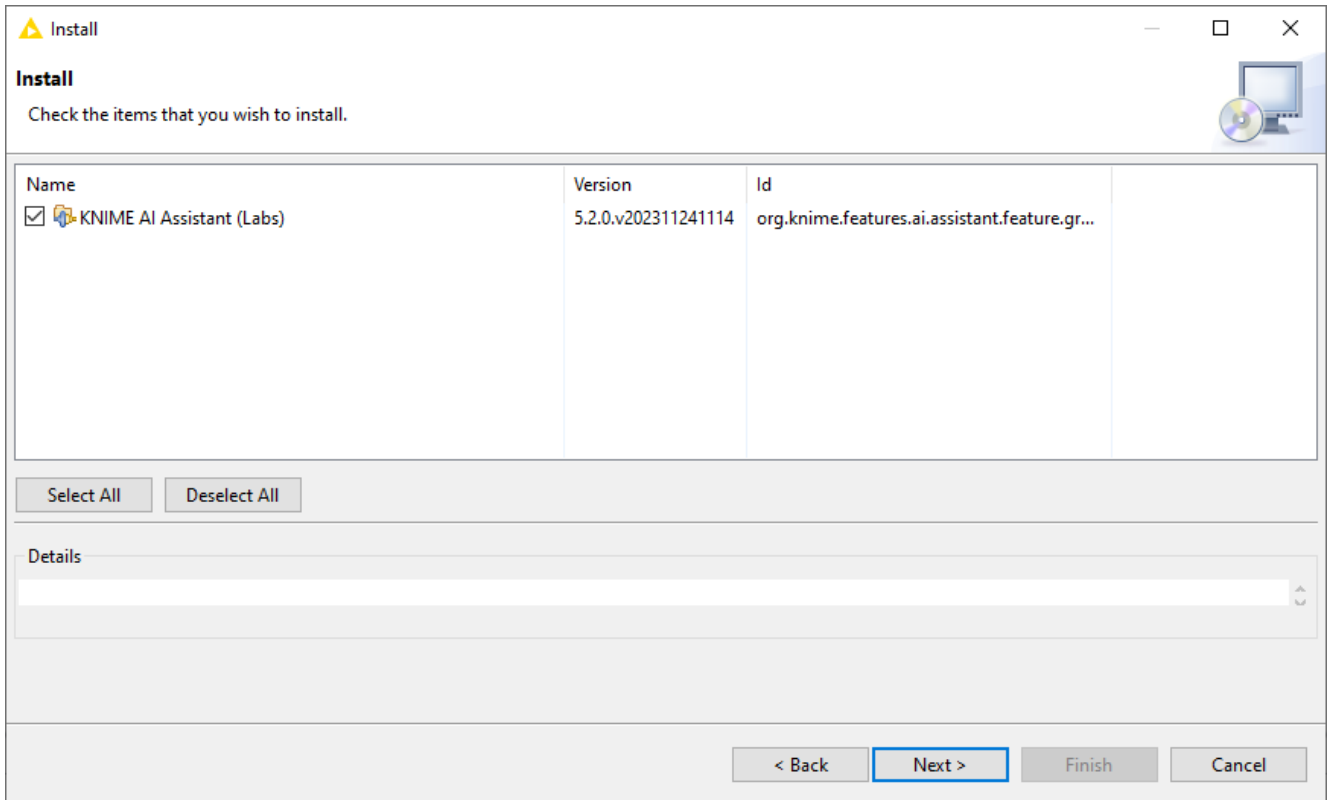


图21. 安装菜单

## 使用

要访问AI助手，请登录KNIME Hub。

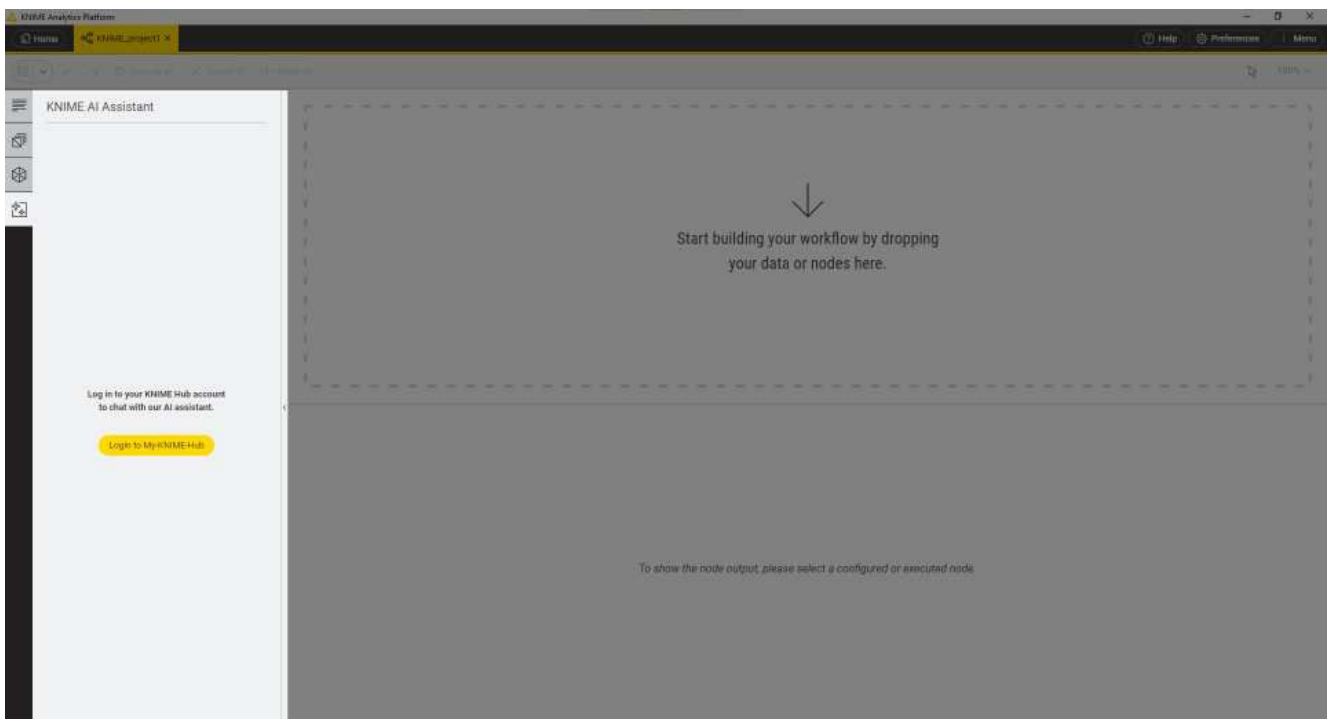


图22. 登录

如果您可以访问配备有AI助手支持的KNIME Business Hub实例，则可以通过AI助手首选项页面选择特定实例并通过AI助手侧边栏登录。

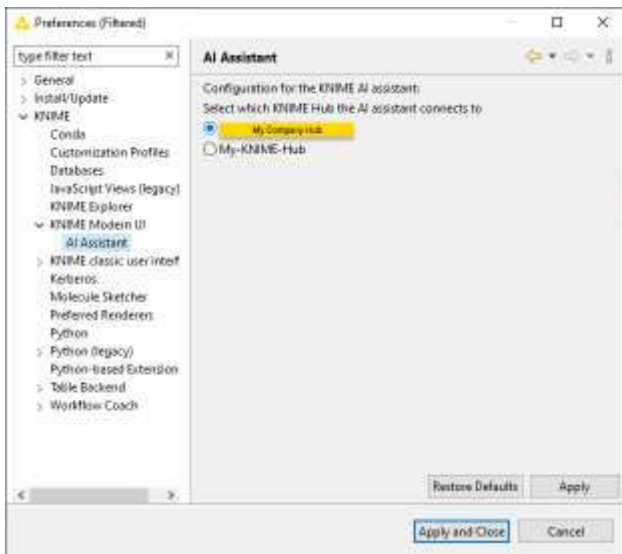


图23. AI助手首选项页面

使用AI助手前，必须先接受免责声明中的条款。

请注意，为了提供我们的服务，KNIME与OpenAI或Microsoft Azure共享数据。这包括所有用户查询以及“构建”模式下所选节点的表格规范，如列名和数据类型，但不包括数据本身。

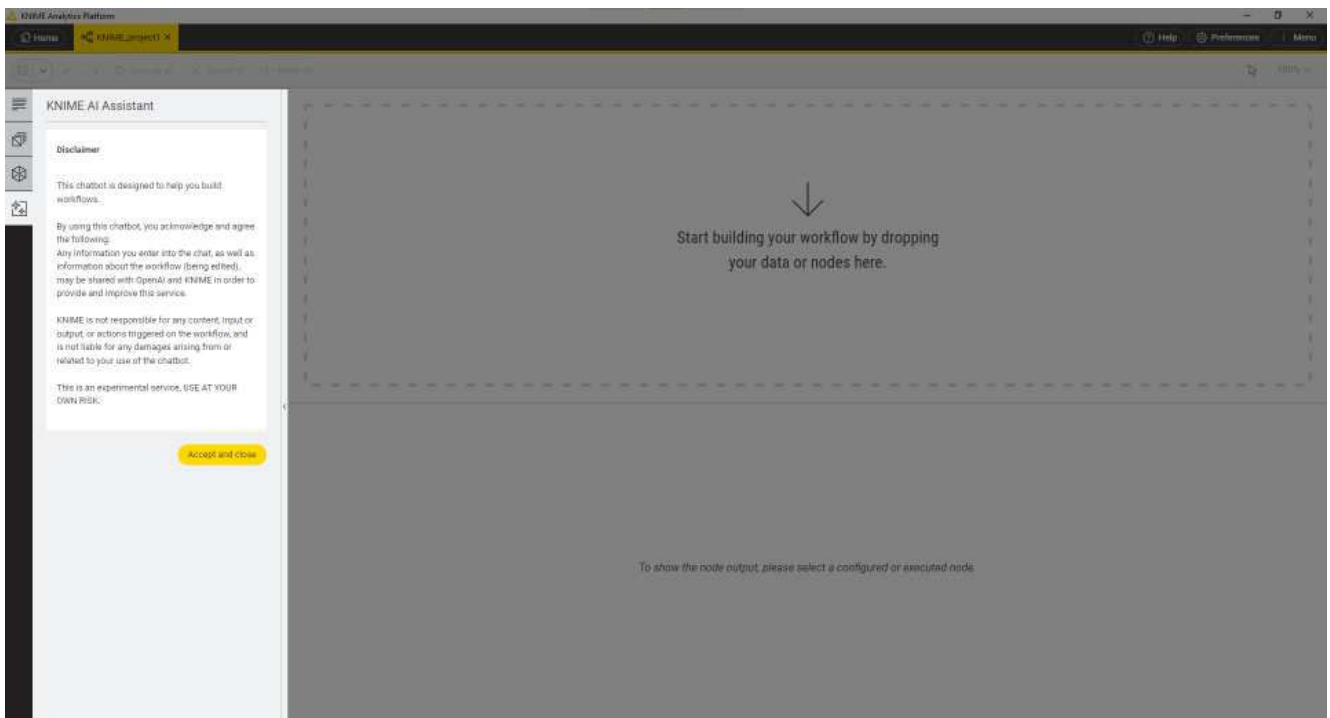


图24. 免责声明

KNIME AI助手提供两种模式：

- 问答模式，和
- 构建模式。

可以通过侧边栏顶部的切换按钮选择这两种模式。

## 问答模式

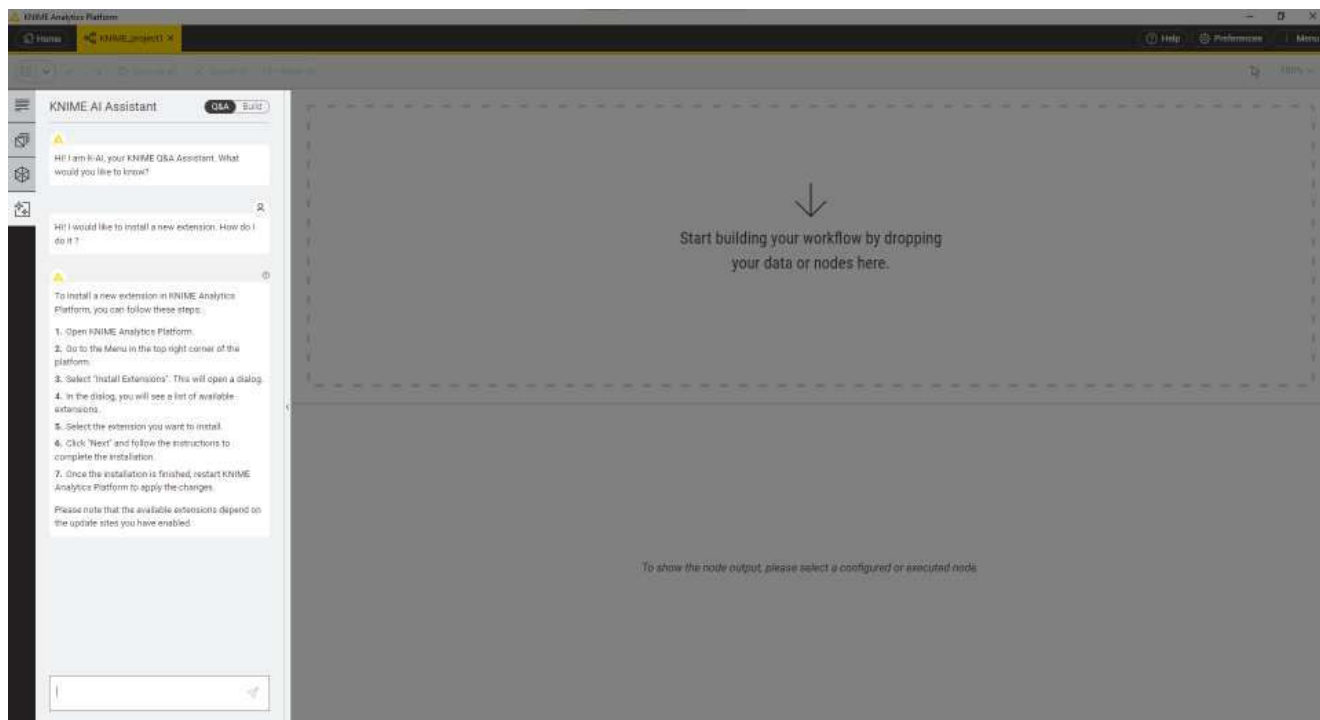


图25. 问答模式

在问答模式下，您可以查询KNIME的功能，包括如何执行特定任务，并获得有用的答案。

这些答案可能会提供有效完成任务所需的节点建议。如果建议的节点已安装，可以直接拖入工作流程中。对于尚未安装的节点，提供了到KNIME Hub的链接。然后，您可以通过从KNIME Hub拖放来安装这些节点。

点击答案顶部的问号，将提供用于生成响应的来源链接。



图26. 问答模式

## 构建模式

构建模式旨在根据查询扩展工作流程。目前构建模式可用的节点集合有限，但将来会添加更多节点。重要的是要注意，在构建模式下，目前无法从头开始初始化工作流程。您需要选择已经提供数据的预先存在的节点，然后根据您的查询动态扩展工作流程。

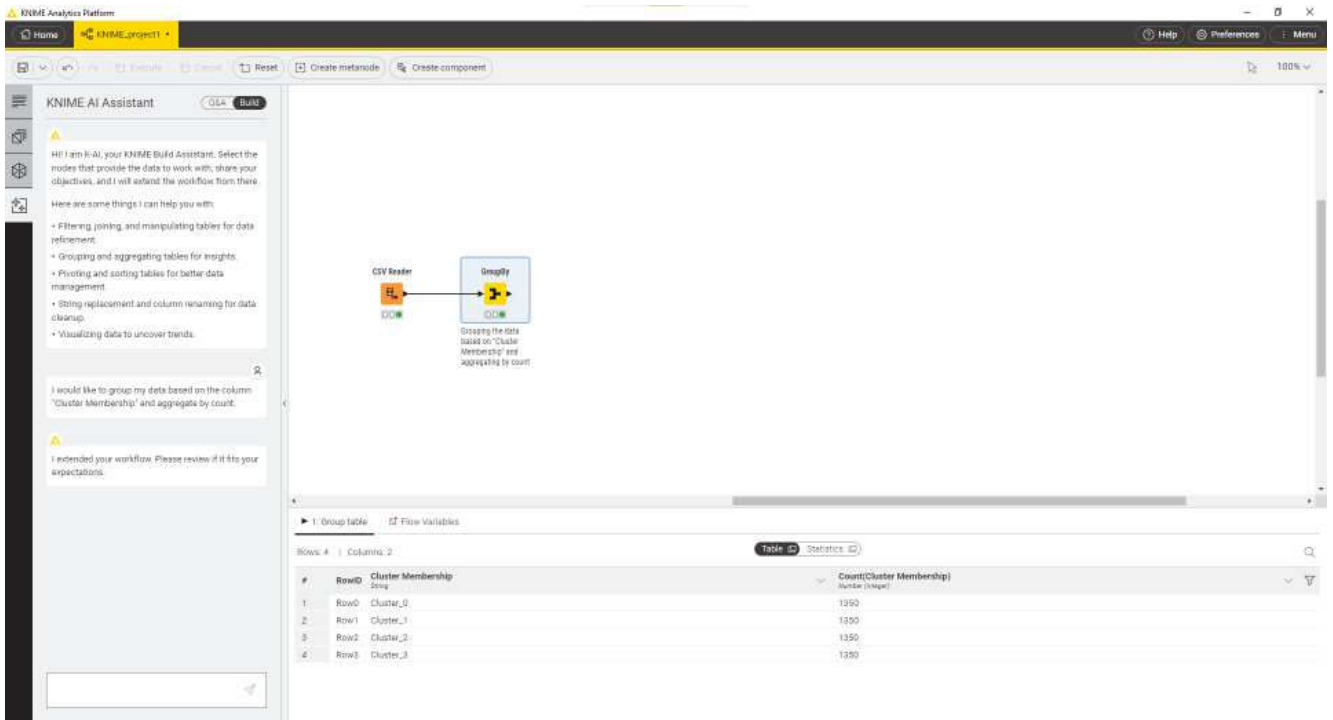


图27. 构建模式

## 节点监视器

节点监视器选项卡位于用户界面底部，如图28所示。它特别适用于检查工作流程中的中间输出表。

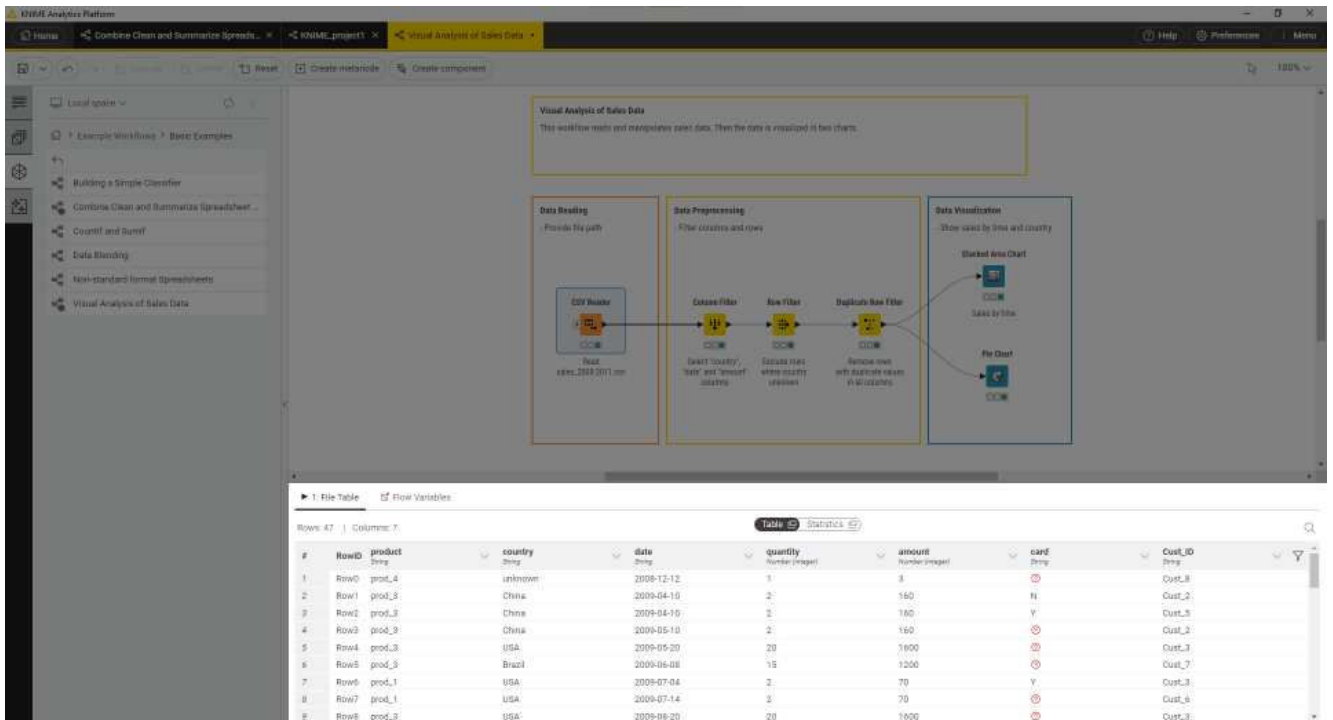


图28. 节点监视器

在这里，您可以选择在活动工作流程中的任何端口上显示流变量或输出数据的预览。

选择活动工作流中所选节点的流变量或输出数据的预览。

切换到统计以查看数据的基本统计信息。

您还可以将表格或统计视图分离并在新窗口中打开。要这样做，请单击相应视图（表格或统计）中的图标。这样可以在工作流中的多个节点上打开多个表格或统计信息。

在KNIME表格部分中，了解更多有关节点监视器中显示的数据表的信息。



## 帮助

通过点击用户界面右上角的帮助按钮，您可以看到多个有用的链接，例如：

- 一些学习资源，如备忘单、入门指南和文档
- KNIME论坛，可向社区询问有关 workflow 构建、技巧和窍门的问题
- KNIME Analytics Platform 的关于页面 - 同时显示当前安装版本和访问安装详细信息
- 关于开源软件组件的额外积分

## 自定义分析平台

### 重置和日志记录

当节点被重置时，节点状态从“已执行”变为“已配置”，节点的输出不再可用。在保存已执行状态的工作流时，也会保存工作流中使用的数据。也就是说，数据集越大，文件大小越大。因此，在保存工作流之前建议先重置工作流，以防数据集可以无限制地访问。

重置工作流只保存节点配置，不保存任何结果。但是，重置节点不会撤消之前执行的操作。在创建、配置和执行工作流期间进行的所有操作都会在 `knime.log` 文件中记录。

点击菜单 > 在文件资源管理器中显示 *KNIME* 日志在用户界面的右上角，以查看 `knime.log` 的位置文件夹。文件 `knime.log` 的大小有限，达到后行将从顶部开始被覆盖。

# 配置KNIME分析平台

## 首选项

随着KNIME Analytics Platform版本5.1的发布，首选项已经重新排列。可以从现代用户界面中打开，点击Analytics Platform右上角的首选项。

在打开的对话框中，显示了一个子类别列表。每个类别都包含一个单独的对话框，用于特定设置，如数据库驱动程序、可用更新站点和外观。

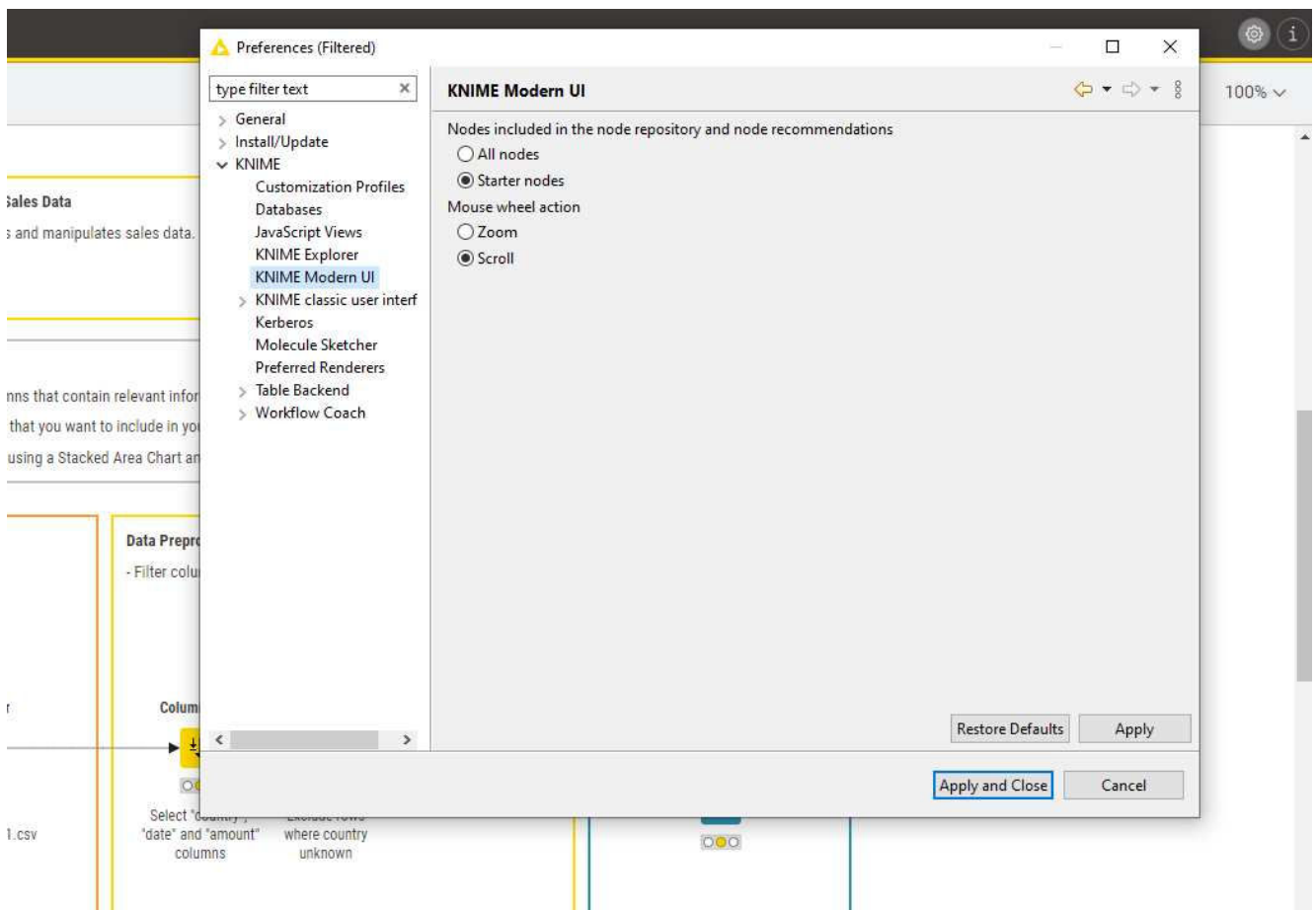


图29。从KNIME Analytics Platform打开KNIME首选项窗口

## KNIME

在子类别列表中选择KNIME，可以定义日志文件的日志级别。默认情况下，它设置为DEB UG。此日志级别可帮助开发人员找到任何意外行为的原因。

直接下方，您可以定义所有节点的最大线程数。工作流的不同分支分配给多个线程，以优化整体执行时间。默认情况下，线程数设置为运行机器上CPU数量的两倍。

在同一个对话框中，您还可以定义临时文件的文件夹。

勾选最后一个选项是的，帮助改进KNIME。同意向我们发送匿名使用数据。此协议激活了工作流教练中社区的节点推荐。

### KNIME现代用户界面

在KNIME现代用户界面类别中，您可以：

- 选择要包含在节点存储库和节点推荐中的节点
- 选择与鼠标滚轮关联的操作。
- 在AI助手下，您还可以选择AI助手连接到的KNIME Hub。

### KNIME经典用户界面

KNIME类别中包含一个子类别 KNIME经典用户界面。在此对话框中，您可以定义控制台视图的日志级别。默认设置为"WARN"，因为更详细的信息仅用于诊断目的。

在使用KNIME Analytics Platform时，您可以选择显示哪些确认对话框。可选择以下选项：

- 重置节点后确认
- 删除节点或连接
- 替换连接
- 保存和执行工作流
- 加载使用夜间构建创建的工作流

在同一个对话框中，您可以定义如果操作需要执行工作流中的前一个节点时会发生什么。您有以下三个选项：

- 自动执行节点
  - 总是拒绝节点执行
-

- 显示对话框以执行或不执行

以下选项允许您定义工作流是否应自动保存以及保存时间间隔，还可以自动更新链接组件和元节点。您还可以定义可视属性，例如工作流注释的边框宽度。

## 表后端

从KNIME Analytics Platform版本4.3开始，引入了新的列后端，以优化KNIME Analytics Platform中主存储器的使用，通过审查底层数据表示来表示表中的单元元素。

KNIME列式表后端扩展通过使用不同的底层数据层（由Apache Arrow支持），基于列式表示来解决这些问题。

可以定义使用的表后端类型：

- 作为所有新工作流的默认设置。打开KNIME首选项并在首选项窗口的左窗格下选择表后端 → KNIME。在这里，您可以选择将列式后端作为新工作流的表后端，如图30所示。

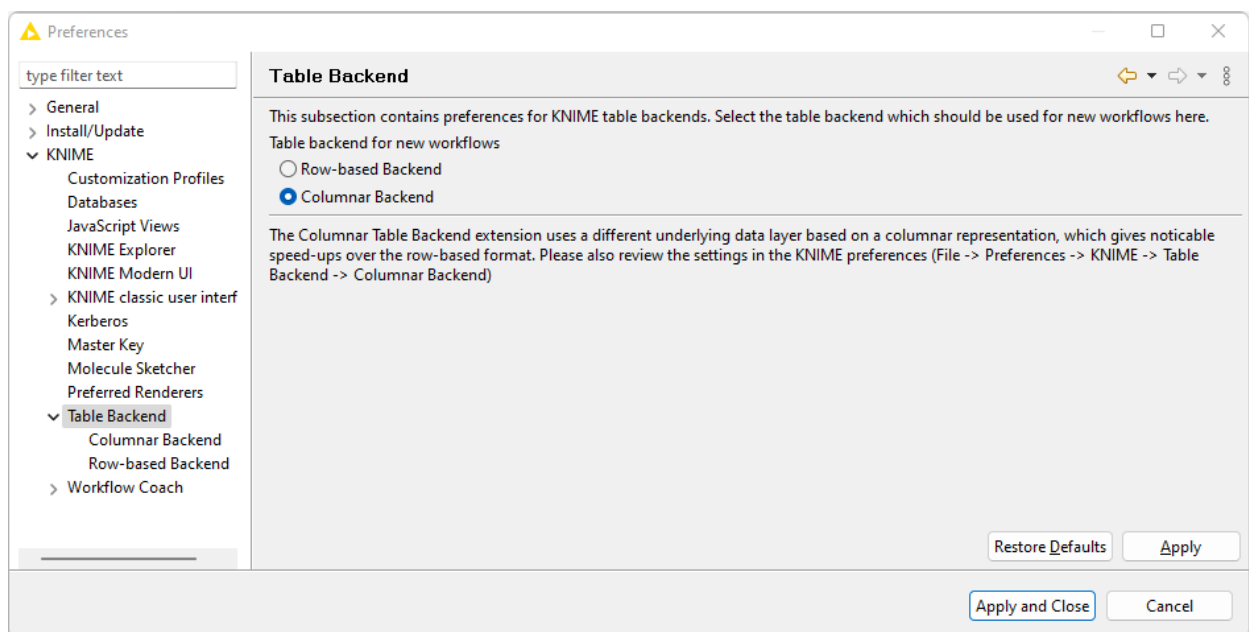


图30。表后端首选项页面。

还可以配置与列式后端的内存使用相关的参数。

转到文件 → 首选项并在首选项窗口的左窗格下选择表后端 → 列式后端，如图31所示。

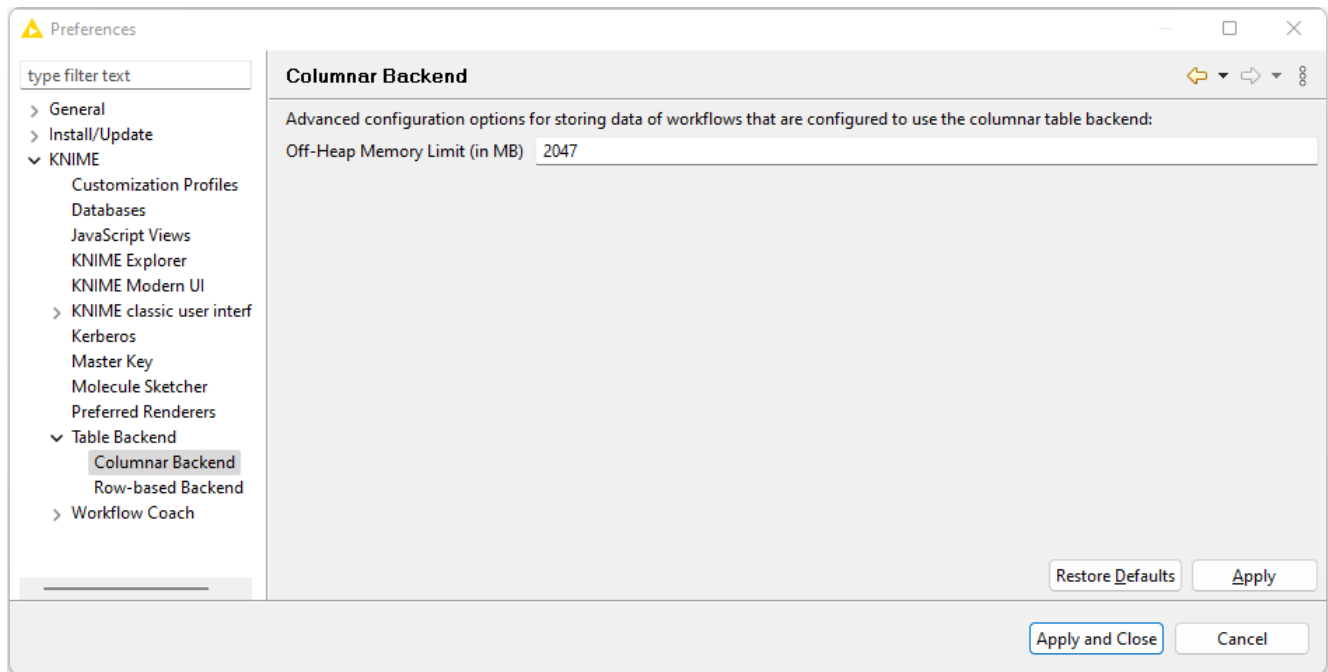


图31。列式后端首选项页面。

请注意，驻留在堆外内存区域中的列式后端缓存除了您在 `knime.ini` 中通过 `-Xmx` 参数分配给 KNIME 的 Java 虚拟机的堆空间之外，还需要一定的内存量。在通过首选项页面更改这些缓存的大小时，请确保不要超过系统的物理内存大小，否则可能会遇到系统不稳定甚至崩溃的情况。



有关列式后端技术背景的更详细解释，请参阅 KNIME 博客上的这篇文章。

**Linux 上的高内存使用：**在某些 Linux 系统上，使用列式后端时，KNIME Analytics Platform 可能会分配比预期更多的系统内存。这是由 JVM 和 `glibc` 本地内存分配器之间的不利交互引起的。有多种方法可以解决这个问题。

- 选项1：减少允许的 `malloc` 区域数量

1. 使用环境变量 `MALLOC_ARENA_MAX` 设置为 1 运行 KNIME Analytics 平台。

- 选项2：使用 `jemalloc`

1. 在您的操作系统上安装 `jemalloc`。对于 Ubuntu：`apt install libjemalloc2`（链接到软件包）。

2. 找到 `jemalloc` 的路径：`$ ldconfig -p | grep jemalloc`。在 Ubuntu 22.04 上，它位于 `lib/x86_64-linux-gnu/libjemalloc.so.2`。

3. 使用环境变量LD\_PRELOAD=<步骤2中的libjemalloc.so路径>启动KNIME Analytics Platform。

• 选项3：使用tcmalloc

1. 在操作系统上安装tcmalloc。对于Ubuntu：apt install google-preftools。

2. 找到tcmalloc的路径：\$ ldconfig -p | grep tcmalloc。在Ubuntu 22.04上，它位于 /lib/x86\_64-linux-gnu/libtcmalloc.so.4。

3. 使用环境变量LD\_PRELOAD=<步骤2中的libtcmalloc.so路径>启动KNIME Analytics Platform。

## 设置knime.ini

在安装KNIME Analytics Platform时，配置选项被设置为默认值。配置选项，即KNIME Analytics Platform使用的选项，从内存设置到某些扩展所需的系统属性都有。

您可以在 knime.ini文件中更改默认设置。 knime.ini文件位于KNIME Analytics Platform的安装文件夹中。



要在MacOS上找到 knime.ini文件，请打开Finder并导航到已安装的应用程序。

接下来，右键单击KNIME应用程序，在菜单中选择显示包内容，并导航到内容，并打开 *Eclipse*。

使用任何纯文本编辑器（如Notepad（Windows），TextEdit（MacOS）或gedit（Linux））编辑knime.ini文件。

在 knime.ini文件中，条目 -Xmx1024m指定KNIME AnalyticsPlatform允许使用的内存量。此值的设置取决于运行机器中可用的内存量。我们建议将其设置为可用内存的大约一半，但此值可以修改和个性化。例如，如果计算机有16GB的内存，则可能将条目设置为 -Xmx8G。

---

除了可用内存外，您还可以在 `knime.ini` 文件中定义许多其他设置。在表1中或配置选项的完整列表中找到一些常见设置的概述。

表格 1. `knime.ini` 文件中的常见配置设置

设置	解释
<code>-Xmx</code> <ul style="list-style-type: none"> <li>• 默认值：1024m</li> <li>• 示例：-Xmx16G</li> </ul>	设置 KNIME Analytics Platform 可用的最大内存量。
<code>-Dknime.compress.io</code> <ul style="list-style-type: none"> <li>• 默认值：SNAPPY</li> <li>• 可能的值：[SNAPPY GZIP NONE]</li> <li>• 示例： -Dknime.compress.io=SNAPPY</li> </ul>	确定在将临时表写入磁盘时使用的压缩算法（如果有）。
<code>-Dorg.knime.container.cellsinmemory</code> <ul style="list-style-type: none"> <li>• 默认值：5,000</li> <li>• 可能的值：0 到 2,147,483,647 之间的任何值</li> <li>• 示例： -Dorg.knime.container.cellsinmemory=100,000</li> </ul>	此设置定义了“小表”的大小。尝试将小表保留在内存中，独立于表缓存策略。通过增加小表的大小，可以限制磁盘交换的次数，但这会减少其他操作可用的内存空间。
<code>-Dknime.layout_editor.browser</code> <ul style="list-style-type: none"> <li>• 默认值版本4.7.2：swt</li> <li>• 可能的值：[cef swt]</li> <li>• 示例： -Dknime.layout_editor.browser=cef</li> </ul>	此设置定义了用于显示布局编辑器的浏览器。

设置	解释
-Dknime.table.cache <ul style="list-style-type: none"> <li>• 默认值: LRU</li> <li>• 可能的值: [LRU SMALL]</li> <li>• 示例: -Dknime.table.cache=SMALL</li> </ul>	确定是否尝试将大表（即不被视为“小”表的表；请参阅设置-Dorg.knime.container.cellsinmemory）缓存在内存中。如果设置为LRU，则大表按最近最少使用（LRU）的顺序或在内存不足时从内存中驱逐。如果设置为SMALL，则大表始终被刷新到磁盘。
-Dknime.url.timeout <ul style="list-style-type: none"> <li>• 默认值: 1,000毫秒</li> <li>• 例子:-Dknime.url.timeout=100</li> </ul>	当尝试连接或从URL读取数据时，此值定义了请求的超时时间。如果读取节点失败，请增加该值。超时时间设置过高可能导致KNIME Analytics Platform中的网站加载缓慢或阻塞对话框。
-Dchromium.block_all_external_requests <ul style="list-style-type: none"> <li>• 默认值: false</li> <li>• 示例: -Dchromium.block_all_external_requests=true</li> </ul>	当设置为true时，此配置设置将阻止Chromium Embedded Framework发出的所有外部请求。
-Dorg.knime.ui.feature.ai_assistant <ul style="list-style-type: none"> <li>• 默认值: true</li> <li>• 示例: -Dorg.knime.ui.feature.ai_assistant=false</li> </ul>	当设置为false时，此配置设置将禁用KNIME AI助手。

## KNIME运行时选项

通过在启动过程中的命令行上传递选项，可以以多种方式配置KNIME的运行行为。由于KNIME基于Eclipse，所有Eclipse运行时选项也适用于KNIME。

KNIME还添加了其他选项，下面将进行描述。



## 命令行参数

下面列出了KNIME处理的命令行参数。它们可以在KNIME安装根目录下的knime.ini文件中永久指定，也可以传递给KNIME可执行文件。请注意，命令行参数必须在系统属性（见下文）之前指定，即在-vmargs参数之前。

请注意，无头KNIME应用程序（如批处理执行器）提供了相当多的命令行参数。这里没有描述它们，但如果您在没有任何参数的情况下调用应用程序，它们将被打印出来。

### **-checkForUpdates**

如果使用此参数，KNIME将在启动时自动检查更新。如果找到已安装功能的新版本，用户将被提示安装它们。安装更新后需要重新启动。

## Java系统属性

下面列出了可以更改KNIME行为的Java系统属性。

它们可以在KNIME安装根目录下的knime.ini文件中永久指定，也可以传递给KNIME可执行文件。请注意，系统属性必须在-vmargs参数之后指定。所需的格式是-DpropertyName=propValue。

## 常规属性

### **org.knime.core.maxThreads=<数量>**

设置KNIME用于执行节点的最大线程数。默认情况下，此数字是核心数的1.5倍。此属性覆盖了KNIME首选项页面中的值。

### **knime.tmpdir=<目录>**

设置临时文件（如数据文件）的默认目录。  
此属性覆盖了首选项页面中的值，并且默认情况下与java.io.tmpdir相同。

---

**knime.synchronous.io=(true|false)**

可用于强制对KNIME表的行进行顺序处理。默认情况下，每个表容器以异步方式在一些（可能是重复使用的）线程中处理其行。默认值为false。将此字段设置为true将指示KNIME始终按顺序和同步方式处理行，在某些情况下可能会更慢。

**knime.async.io.cachesize=<数字>**

设置非顺序和异步处理行的批处理大小（参见knime.synchronous.io）。它指定单个容器线程处理的数据行数。缓冲区越大，同步开销越小，但内存需求越大。如果行按顺序处理，则此属性无效。默认值为10。

**knime.domain.valuecount=<数字>**

在向表中添加行时保留的名义值数量。这只是默认值，可能会被各个节点实现覆盖。如果未指定值，则使用默认值60。

**org.knime.container.threads.total=<数字>**

设置可用于写入KNIME本机输出表的最大线程数。默认情况下，此数字等于JVM可用的处理器数量。注意：此值必须大于0。

**org.knime.container.threads.instance=<数字>**

设置可以用于写入单个KNIME本地输出表的最大线程数。默认情况下，此数字等于JVM可用的处理器数量。注意：此值必须大于0且不能大于org.knime.container.threads.total。

**knime.discourage.gc=(true|false)**

如果设置为true，则不鼓励KNIME触发全停顿垃圾回收。请注意，（a）单个节点可以忽略此设置，（b）垃圾回收器可能会独立决定是否需要进行全停顿垃圾回收。默认情况下设置为true。

**org.knime.container.minspace.temp=<数字>**

Java属性，用于指定需要可用的最小空闲磁盘空间（以MB为单位）。

如果可用空间较少，则不会创建更多的表文件和blob（导致异常）。

**knime.columnar.chunksize=<数字>**

列式表后端将表水平分成批次，并将这些批次垂直分成列块。此属性控制这些块的初始大小，从而控制每个批次的行数。块是必须材料化以访问单个值的最小单位。因此，更改此值可能会影响内存占用和整体性能。除非有充分的理由，否则不要更改此值。默认值为28,000。

**knime.columnar.reservedmemorymb=<数字>**

列式表后端将表数据缓存在堆外。为此，除了JVM的堆内存之外，还需要内存，其大小通过-Xmx参数控制。如果在首选项中未设置显式缓存大小，则默认可用于缓存的内存计算如下：总物理内存减去保留内存减去1.25倍的堆内存。此方程中的保留内存大小（以MB为单位）可以通过此属性进行配置。默认值为4,096。

**knime.columnar.verbose=(true|false)**

将此属性设置为true将激活列式表后端的详细调试日志记录。

-

**knime.disable.rowid.duplicatecheck=(true|false)**

启用/禁用表格上的行ID重复检查。在KNIME中，表格应具有唯一的ID，通过使用重复检查器来确保唯一性。

此属性将禁用此检查。

警告：用户不应更改此属性。

**knime.disable.vmlfilelock=(true|false)**

启用/禁用工作流程锁定。从KNIME 2.4开始，打开工作流程时将锁定工作流程；此属性将禁用锁定（允许多个实例打开相同的工作流程）。

警告：用户不应更改此属性。

**knime.database.timeout=<数字>**

设置尝试与数据库建立连接的超时时间（以秒为单位）。  
默认值为15秒。

**knime.database.fetchsize=<数字>**

设置从数据库检索数据的获取大小。  
默认值取决于所使用的JDBC驱动程序。

**knime.database.batch\_write\_size=<数字>**

设置将数据行写入数据库的批量写入大小。  
默认值为1，即一次一行。

**knime.database.enable.concurrency=(是|否)**

用于开启/关闭数据库连接访问（仅适用于相同的数据库连接）。

默认值为true，即所有数据库访问都基于单个连接进行同步；false表示关闭，即访问不进行同步，可能导致数据库错误。

**knime.logfile.maxsize=<数字>[兆字节]**

允许更改最大日志文件大小（默认为10 MB）。值必须是整数，可能后跟"m"或"k"以表示给定值为兆字节或千字节。

**knime.settings.passwords.forbidden=(是|否)**

如果为true，则使用密码作为配置的节点（例如DB连接或SendEmail）将不会将密码存储为工作流程的一部分。相反，将存储一个空值，这将导致节点的配置在从磁盘恢复工作流程后不正确（但有效）。默认值为false。

**knime.repository.non-instant-search=(true|false)**

允许禁用节点库搜索中的实时更新。

-

**knime.macosx.dialogworkaround=(true|false)**

允许禁用在MacOSX下打开节点对话框时出现冻结的解决方法。

-

**knime.data.bitvector.maxDisplayBits=<number>**

设置在位向量的字符串表示中显示的最大位数。

-

**knime.xml.disable\_external\_entities=(true|false)**

如果设置为true，则所有解析XML文件的节点将不会读取通过DTD定义的外部实体。

这通常只在作为服务器上的执行程序运行时有用，并且您想要防止XXE攻击时使用。

## 插件相关属性

这些属性仅影响某些插件，并且仅在安装了这些插件时适用。

**org.knime.cmlminblobsize=<number>[mMkK]**

允许更改以字节（或千字节或兆字节）为单位的CML分子在存储在blob单元之前必须具有的最小大小。否则，它将内联存储。后者速度稍快，但需要更多内存。默认值为8kB。

**org.knime.ctabminblobsize=<数字>[毫米克千]**

允许更改Ctab分子在存储为blob单元之前必须具有的最小大小（以字节、千字节或兆字节为单位）的设置。否则，它将内联存储。后者速度稍快，但需要更多内存。默认值为8kB。

**org.knime.mol2minblobsize=<数字>[毫米克千]**

允许更改Mol2分子在存储为blob单元之前必须具有的最小大小（以字节、千字节或兆字节为单位）的设置。否则，它将内联存储。后者速度稍快，但需要更多内存。默认值为8kB。

**org.knime.molminblobsize=<数字>[毫米克千]**

允许更改Mol分子在存储为blob单元之前必须具有的最小大小（以字节、千字节或兆字节为单位）的设置。否则，它将内联存储。后者速度稍快，但需要更多内存。默认值为8kB。

**org.knime.rxnminblobsize=<数字>[毫米克]**

允许更改在将Rxn分子存储在blob单元之前必须具有的最小大小（以字节、千字节或兆字节为单位）。否则，它将内联存储。后者速度稍快，但需要更多内存。默认值为8kB。

**org.knime.sdfminblobsize=<数字>[毫米克]**

允许更改在将SDF分子存储在blob单元之前必须具有的最小大小（以字节、千字节或兆字节为单位）。否则，它将内联存储。后者速度稍快，但需要更多内存。默认值为8kB。

# KNIME表格

## 数据表

节点的非常常见的输入和输出端口是数据输入端口和数据输出端口，对应于图32中的黑色三角形。

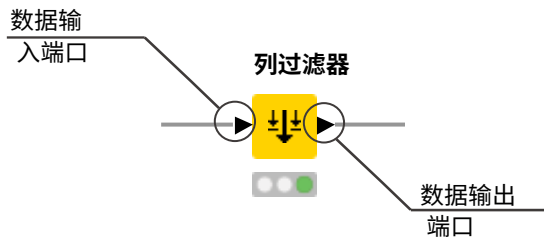


图32. 数据输入和输出端口

数据表由列和行组织，它包含一定数量的等长行。每列中的元素必须具有相同的数据类型。

图33所示的数据表是由CSV Reader节点生成的，它是许多具有黑色三角形输出端口的节点之一。要打开表格，请单击该节点。如果尚未执行该节点，请执行该节点。表格将显示在节点监视器中。

输出表格具有行号、唯一的RowID和列标题。RowID由读取器节点自动创建，但也可以手动定义。因此，RowID和列标题可以用于标识表格中的每个数据单元。数据中的缺失值以红色问号在圆圈中显示。

在节点监视器的顶部，您可以通过选项卡和流变量选项卡选择要查看的输出端口，流变量选项卡显示节点输出中可用的流变量及其当前值。下一行将指示表格的维度，即该特定输出端口中有多少行和多少列。在这里，您还可以使用切换按钮切换到统计信息。此选项卡显示表格的元信息，如列名、列类型和其他一些统计数据。

► 1: File Table Flow Variables

Rows: 47 | Columns: 7 Table **Statistics**

#	Row...	product String	country String	date String	quantity Number (integer)	amount Number (integer)	card String	Cust_ID String
1	Row0	prod_4	unknown	2008-12-12	1	3	⊖	Cust_8
2	Row1	prod_3	China	2009-04-10	2	160	N	Cust_2
3	Row2	prod_3	China	2009-04-10	2	160	Y	Cust_5
4	Row3	prod_3	China	2009-05-10	2	160	⊖	Cust_2
5	Row4	prod_3	USA	2009-05-20	20	1600	⊖	Cust_3
6	Row5	prod_3	Brazil	2009-06-08	15	1200	⊖	Cust_7
7	Row6	prod_1	USA	2009-07-04	2	70	Y	Cust_3
8	Row7	prod_1	USA	2009-07-14	2	70	⊖	Cust_6
9	Row8	prod_3	USA	2009-08-20	20	1600	⊖	Cust_3
10	Row9	prod_2	Germany	2009-11-02	15	600	⊖	Cust_1
11	Row10	prod_2	Germany	2009-11-22	15	600	N	Cust_1
12	Row11	prod_1	Germany	2009-12-02	1	35	Y	Cust_1
13	Row12	prod_1	China	2009-12-12	1	35	Y	Cust_2
14	Row13	prod_3	USA	2010-01-03	20	1600	⊖	Cust_3
15	Row14	prod_1	Germany	2010-01-10	1	35	N	Cust_1
16	Row15	prod_3	Germany	2010-01-13	1	80	⊖	Cust_4
17	Row16	prod_2	Germany	2010-01-15	25	1000	⊖	Cust_1
18	Row17	prod_2	USA	2010-01-20	2	80	⊖	Cust_6
19	Row18	prod_2	USA	2010-02-12	6	240	Y	Cust_6
20	Row19	prod_2	USA	2010-02-22	6	240	⊖	Cust_6
21	Row20	prod_2	Brazil	2010-03-11	6	240	N	Cust_7
22	Row21	prod_3	China	2010-03-12	1	80	⊖	Cust_5

图33. 在KNIME Analytics平台中的数据输出

## 列类型

KNIME Analytics平台中的基本数据类型是整数，双精度和字符串，以及其他支持的数据类型，如长整型，布尔值，JSON，URI，文档，日期和时间，位向量，图像和 Blob。KNIME Analytics平台还支持自定义数据类型，例如分子的表示。

在输出表中切换到统计视图，可以看到数据表中列的数据类型，如图34所示。对于数值，只显示数据中的值范围。对于字符串值，显示数据中出现的不同值。

► 1: File Table Flow Variables

Rows: 7 | Columns: 14 Table **Statistics**

Name	Type	# Missing val...	# Unique valu...	Minimum	Maximum	25% Quantile	50% Quantile ...	75% Quantile	Mean	Mean Absolu...	Standard Dev...	Sum	10 most com...
product	String	0	4	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	prod_1 (16, 34...
country	String	0	5	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	Germany (16; 3...
date	String	0	46	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	2009-04-10 (2; ...
quantity	Number (integ...	0	11	1	25	2	5	10	6.617	5.171	6.357	311	1 (11; 23.4%)...
amount	Number (integ...	0	18	3	1,600	70	175	400	346.234	291.873	416.855	16,273	160 (6; 12.77%...
card	String	24	2	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	Y (15; 65.22%)...
Cust_ID	String	0	9	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	Cust_1 (9; 19.1...

图34. 在"Spec"选项卡中的数据类型和数据域

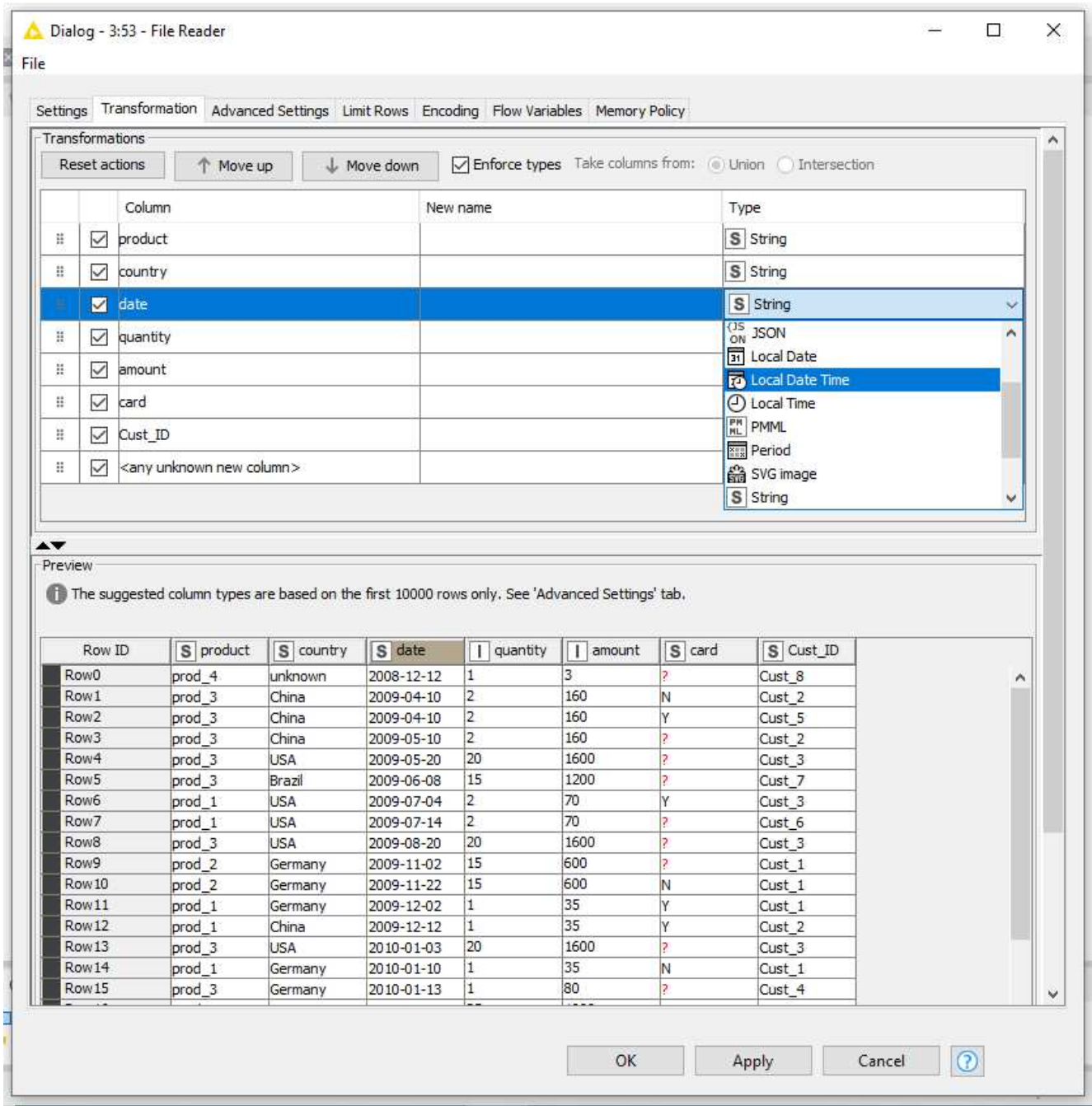
KNIME Analytics平台中的读取器节点根据内容的解释为每个列分配一个数据类型。如果读取器节点无法识别列的正确数据类型，则可以在之后进行更正。有可用的节点来转换数据类型。例如：字符串转数字，数字转字符串，双精度转整数，字符串转日期和时间，字符串转JSON，字符串转URI。

许多特殊的数据类型被读取器节点识别为字符串。要转换这些字符串列为正确的数据类型，请使用列类型自动转换节点。



当您使用文件读取节点读取文件时，可以通过节点配置对列类型进行直接转换。要这样做，请转到配置对话框中的转换选项卡，并更改所需列的类型，如图35所示。

图35。在文件读取节点中更改列类型



通过节点配置对话框直接转换列类型 通过节点配置对话框直接转换列类型

## 排序

表格视图输出中的行可以通过点击悬停在列名上方的向上（升序）和向下（降序）箭头来按照一列的值进行排序。请注意，此排序仅影响当前的输出视图，对节点输出没有影响。

要永久地对输出表格进行排序，请使用排序节点。使用列重新排序节点来重新排序列。

## 列渲染

在表格视图输出中，您还可以更改数字值在数据表中的显示方式。例如，可以将数字值显示为百分比，完整精度，或者用灰度或条形图替换数字。要查看列的这些和其他渲染选项，请点击列标题中的小三角图标，并选择所需的可用渲染器，如图36所示。请注意，这些更改是临时的，对节点输出没有影响。

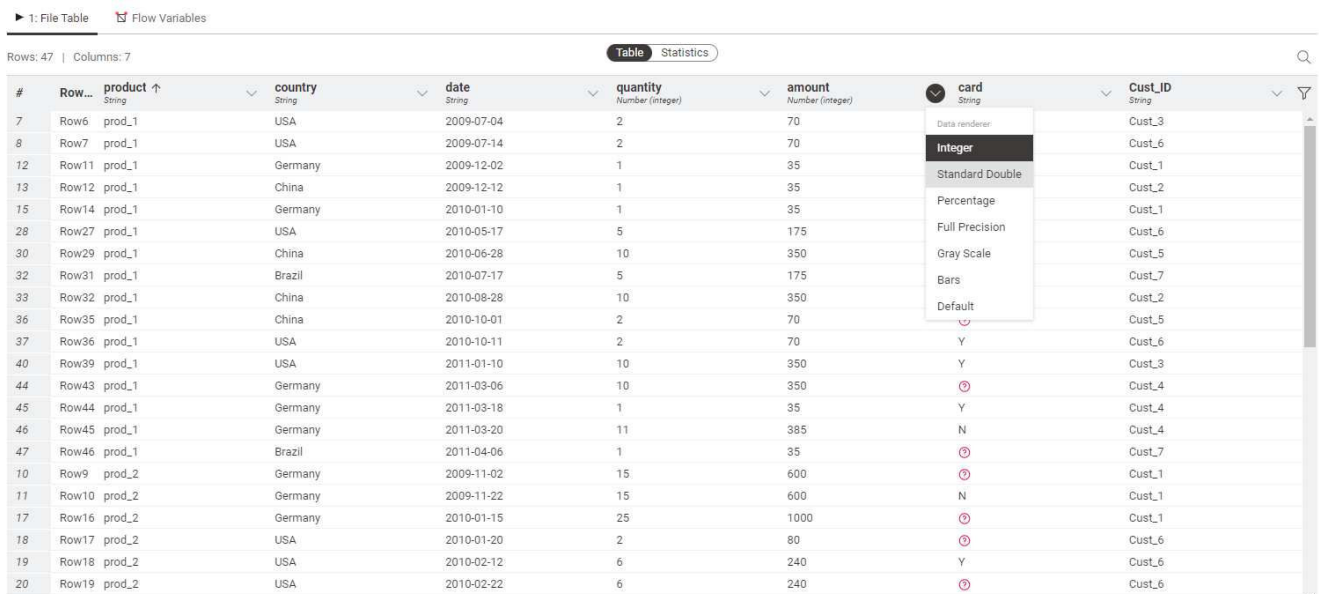


Figure 36 shows a screenshot of the KNIME Analytics Platform interface. The main window displays a data table with 47 rows and 7 columns. The columns are: #, Row..., product, country, date, quantity, amount, card, and Cust\_ID. The 'card' column is currently set to 'Data String'. A dropdown menu is open for the 'card' column, showing various data renderers: Data String, Integer, Standard Double, Percentage, Full Precision, Gray Scale, Bars, and Default. The 'Integer' option is selected. The table data includes columns for product, country, date, quantity, amount, and Cust\_ID.

#	Row...	product	country	date	quantity	amount	card	Cust_ID
7	Row6	prod_1	USA	2009-07-04	2	70		Cust_3
8	Row7	prod_1	USA	2009-07-14	2	70		Cust_6
12	Row11	prod_1	Germany	2009-12-02	1	35		Cust_1
13	Row12	prod_1	China	2009-12-12	1	35		Cust_2
15	Row14	prod_1	Germany	2010-01-10	1	35		Cust_1
28	Row27	prod_1	USA	2010-05-17	5	175		Cust_6
30	Row29	prod_1	China	2010-06-28	10	350		Cust_5
32	Row31	prod_1	Brazil	2010-07-17	5	175		Cust_7
33	Row32	prod_1	China	2010-08-28	10	350		Cust_2
36	Row35	prod_1	China	2010-10-01	2	70		Cust_5
37	Row36	prod_1	USA	2010-10-11	2	70	Y	Cust_6
40	Row39	prod_1	USA	2011-01-10	10	350	Y	Cust_3
44	Row43	prod_1	Germany	2011-03-06	10	350		Cust_4
45	Row44	prod_1	Germany	2011-03-18	1	35	Y	Cust_4
46	Row45	prod_1	Germany	2011-03-20	11	385	N	Cust_4
47	Row46	prod_1	Brazil	2011-04-06	1	35		Cust_7
10	Row9	prod_2	Germany	2009-11-02	15	600		Cust_1
11	Row10	prod_2	Germany	2009-11-22	15	600	N	Cust_1
17	Row16	prod_2	Germany	2010-01-15	25	1000		Cust_1
18	Row17	prod_2	USA	2010-01-20	2	80		Cust_6
19	Row18	prod_2	USA	2010-02-12	6	240	Y	Cust_6
20	Row19	prod_2	USA	2010-02-22	6	240		Cust_6

图36. 以表格视图呈现数据

## 表格存储

当执行时，许多KNIME节点会生成并提供对其输出端口的表格数据访问。这些表格可能是小的或大的，因此可能适合执行机器的主内存中，也可能不适合。可以通过几种选项来配置哪些表格在内存中保留以及何时以及如何将表格写入磁盘。这些选项在本节中概述。

### 内存缓存

KNIME Analytics Platform区分小表格和大表格。当表格由最多（超过）5000个单元格组成时，将认为表格是小的（大的）。可以通过knime.ini文件中的-Dorg.knime.container.cellsinmemory参数来调整5000个单元格的阈值。KNIME Analytics Platform始终尝试将小表格保存在内存中，仅在内存不足时将其刷新到磁盘。

此外，只要有足够的内存，KNIME Analytics Platform会尝试将最近使用的大表格保存在内存中。然而，它会异步地将这些表格写入磁盘，以便在它们长时间未被访问或内存不足时可以从内存中删除。您可以配置特定节点的内存消耗，使其永远不会尝试将其表格保存在内存中，而是在执行时将其写入磁盘。如果您知道一个节点将生成一个无法保存在内存中的表格，或者如果您想减少节点的内存占用，这将非常有帮助。

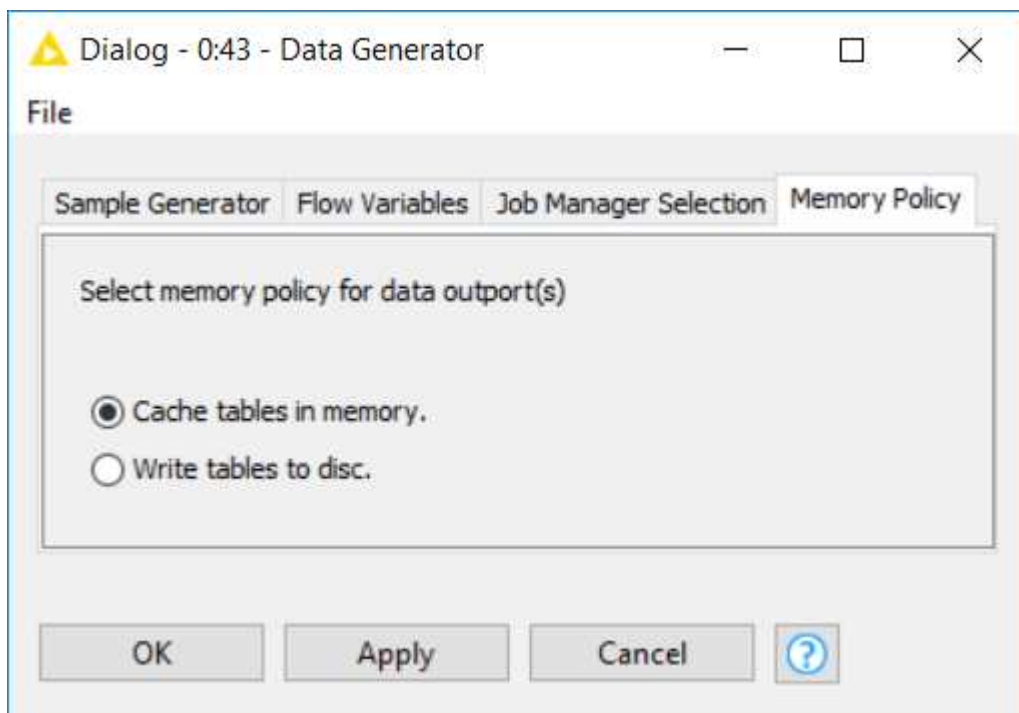


图37. 配置节点的内存策略