

KNIME Workbench Guide

KNIME AG, Zurich, Switzerland
Version 4.2 (last updated on 2020-10-27)



Table of Contents

Workspaces	1
KNIME Workbench	2
Welcome page	4
Workflow editor & nodes	5
KNIME Explorer	13
Workflow Coach	28
Node repository	30
KNIME Hub Search	31
Description	32
Node Monitor	32
Outline	33
Console	33
Customizing the KNIME Workbench	34
Reset and logging	34
Show heap status	34
Configuring KNIME Analytics Platform	35
Preferences	35
Setting up knime.ini	36
KNIME tables	40
Data table	40
Column types	41
Sorting	44
Column rendering	44
Table storage	46

This guide describes the first steps to take after starting KNIME Analytics Platform and points you to the resources available in the KNIME Workbench for building workflows. It also explains how to customize the workbench and configure KNIME Analytics Platform to best suit specific needs. In the last part of this guide we introduce data tables.

Workspaces

When you start KNIME Analytics Platform, the KNIME Analytics Platform launcher window appears and you are asked to define the KNIME workspace, as shown in [Figure 1](#).



The KNIME workspace is a folder on the local computer to store KNIME workflows, node settings, and data produced by the workflow.

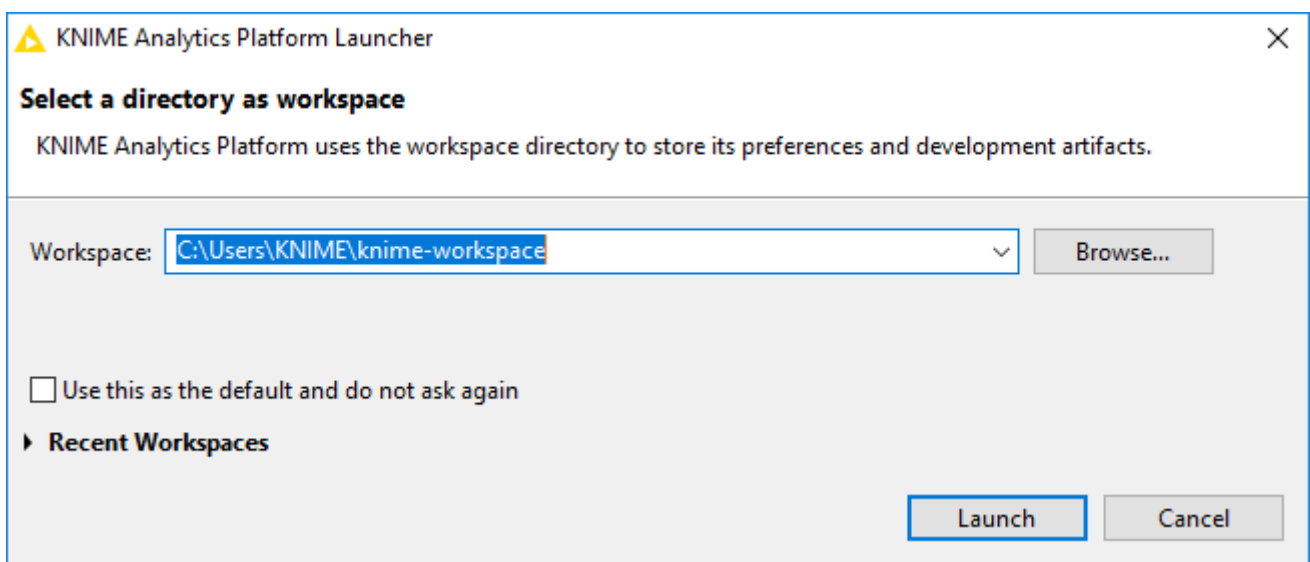


Figure 1. KNIME Analytics Platform launcher

The workflows and data stored in the workspace are available through the KNIME Explorer in the upper left corner of the KNIME Workbench.

KNIME Workbench

After selecting a workspace for the current project, click *Launch*. The KNIME Analytics Platform user interface - the KNIME Workbench - opens.

It is typically organized as shown in **Figure 2**.

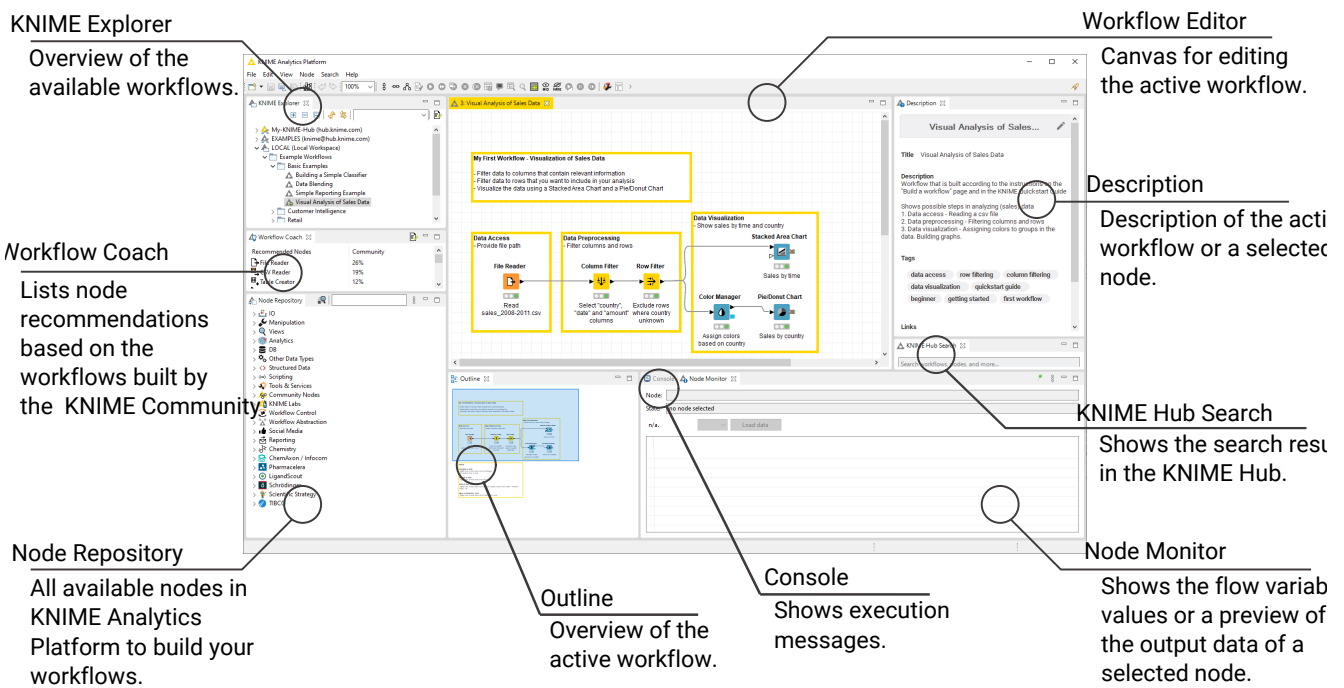


Figure 2. KNIME Workbench

In the next few sections we explain the functionality of these components of the workbench:

- [Welcome Page](#)
- [Workflow Editor & nodes](#)
- [KNIME Explorer](#)
- [Workflow Coach](#)
- [Node Repository](#)
- [KNIME Hub Search](#)
- [Description](#)
- [Node Monitor](#)
- [Outline](#)

- **Console**

Welcome page

The welcome page shown in [Figure 3](#) is located in the middle of the KNIME Workbench.

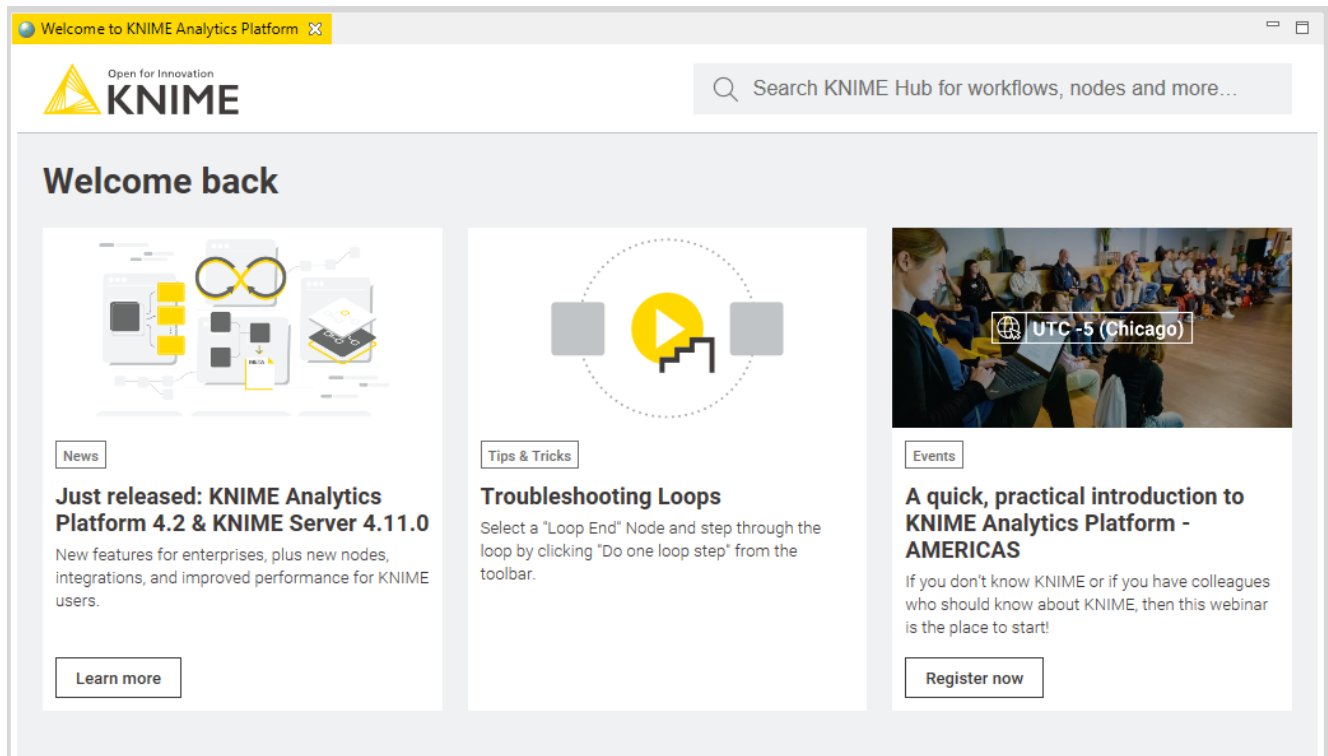


Figure 3. Welcome page

This page links to information, for example available updates and the latest KNIME news, upcoming events, and tips and tricks.

After closing the welcome page, if no previously created workflows are available, you need to create an empty workflow editor, as explained in the next section.

Workflow editor & nodes

The workflow editor is where workflows are assembled. Workflows are made up of individual tasks, represented by nodes.

Create a new workflow editor going to *File* → *New...* and selecting the *New KNIME Workflow* option in the window that opens. Then click *Next* and give the new workflow a name in the field next to *Name of the workflow to create:* and click *Finish*. Other options are available as explained in the [Building workflows section](#).

In the new empty workflow editor, create a workflow by dragging nodes from the [node repository](#) to the workflow editor, then connecting, configuring, and executing them.

Nodes

In KNIME Analytics Platform, individual tasks are represented by nodes. Nodes can perform all sorts of tasks, including reading/writing files, transforming data, training models, creating visualizations, and so on.

Facts about nodes

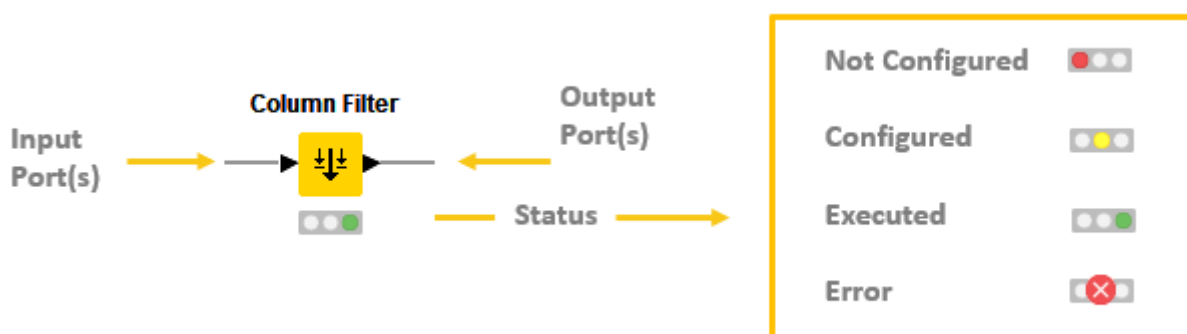


Figure 4. Node ports and node status

- Each node is displayed as a colored box with input and output ports, as well as a status, as shown in [Figure 4](#)
- The input port(s) hold the data that the node processes, and the output port(s) hold the resulting datasets of the operation
- The data is transferred over a connection from the output port of one to the input port of another node.



For simplicity we refer to data when we refer to node input and output ports, but nodes can also have input and output ports that hold a model, a database query, or another type explained in [Node Ports](#).

Changing the status of a node

The status of a node can be changed, either configuring, executing, or resetting it. All these options can be found in the context menu of a node shown in [Figure 5](#).

Open the context menu by right clicking a node. From the context menu it is also possible to open output tables and views, as well as copy nodes, along with some more advanced node options.

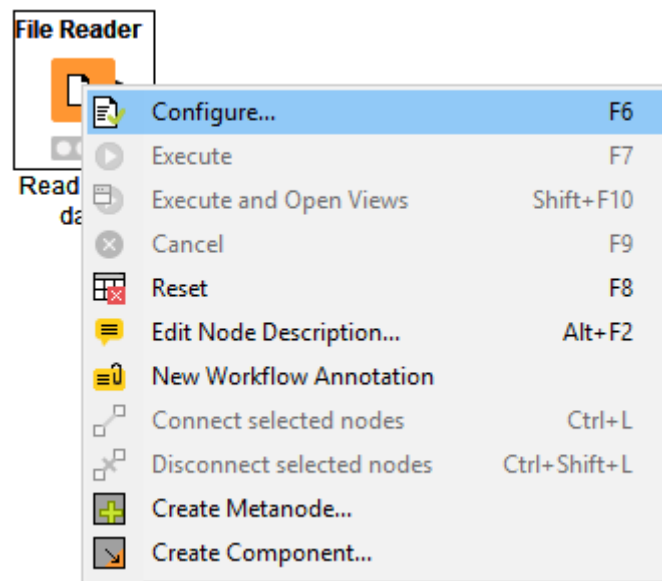


Figure 5. Context menu of a node

Identifying the node status

The traffic light below each node shows the status of the node. When a node is configured, the traffic light changes from red to yellow, i.e. from "not configured" to "configured".

When a new node is first added to the [workflow editor](#), its status is "not configured" - shown by the red traffic light below the node.

Configuring the node

The node can be configured by adjusting the settings in its configuration dialog.

Open the configuration dialog of a node by either:

- Double clicking the node
- Right clicking a node and selecting *Configure...* in the context menu
- Or, selecting the node and pressing F6

In addition to the task specific settings, each node configuration dialog has a:

- "Memory Policy"-tab: here it is possible to define whether tables are attempted to be kept in memory, or if all tables are written to disk (see the section on [In-Memory Caching](#) for details).
- "Flow Variables"-tab: flow variables are explained in the [Flow Control Guide](#)

Executing the node

Some nodes have the status "configured" already when they are created. These nodes are executable without adjusting any of the default settings.

Execute a node by either:

- Right clicking the node and selecting *Execute*
- Or, selecting the node and pressing F7

If execution is successful, the node status becomes "executed", which corresponds to a green traffic light. If the execution fails, an error sign will be shown on the traffic light, and the node settings and inputs will have to be adjusted as necessary.

Right click the node and select one of the last options in the menu to inspect the outputs, such as data tables, and views of an executed node. If the node produces an (interactive) view in its output, like all JavaScript based nodes, select *(Interactive) View: ...* from the context menu to open it.

Canceling execution of the node

To cancel the execution of a node, right click it and select *Cancel* or select it and press F9.

Resetting the node

To reset a node, right click it and select *Reset* or select it and press F8.



Resetting a node also resets all of its subsequent nodes in the workflow. Now, the status of the node(s) turns from "executed" into "configured", the nodes' outputs are cleared.

Node ports

A node may have multiple input ports and multiple output ports. A collection of interconnected nodes, using the input ports on the left and output ports on the right, constitutes a workflow. The input ports consume the data from the output ports of the predecessor nodes, and the output ports provide data to the successor nodes in the workflow.

Besides **data tables**, input and output ports can provide other types of inputs and outputs. For each type the pair of input and output port looks different, as shown in **Figure 6**.

An output port can only be connected to an input port of the same type - data to data, model to model, and so on.

Some input ports can be empty, like the data input port of the Decision Tree View node in **Figure 6**. This means that the input is optional, and the node can be executed without the input. The mandatory inputs, shown by filled input ports, have to be provided to execute the node.

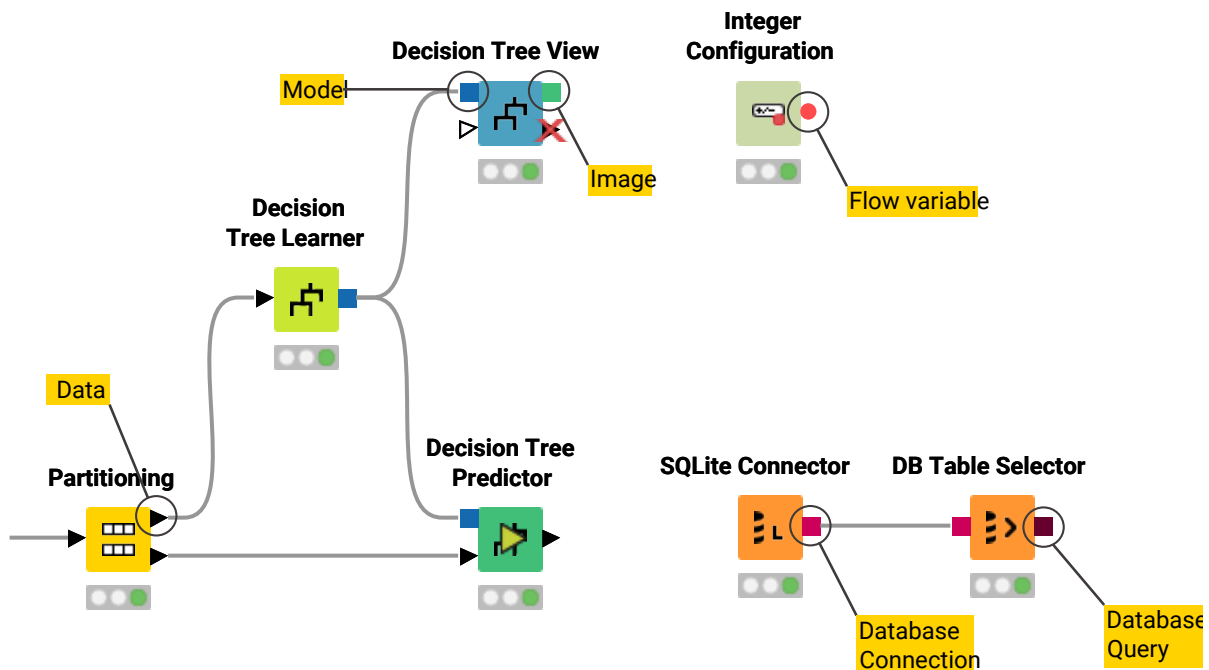


Figure 6. Common port types

A tooltip gives a short explanation of the input and output ports. If the node is executed, the dimensions of the outgoing data are shown in its data output port. A more detailed explanation of the input and output ports is in the node description.

How to select, move, copy, and replace nodes in a workflow

Nodes can be moved into the workflow editor by dragging and dropping them. To copy nodes

between workflows, select the chosen nodes, right click the selection, and select *Copy* in the menu. In the destination workflow, right click the workflow editor, and select *Paste* in the menu.

To select a node in the workflow editor, click it once, and it will be surrounded by a border. To select multiple nodes, either press "Ctrl" and select nodes by mouse click, or draw a rectangle over the nodes with the mouse.

Replace a node by dragging a new node onto an existing node. Now the existing node will be covered with a colored box with an arrow and boxes inside as shown in [Figure 7](#). Releasing the mouse replaces the node.

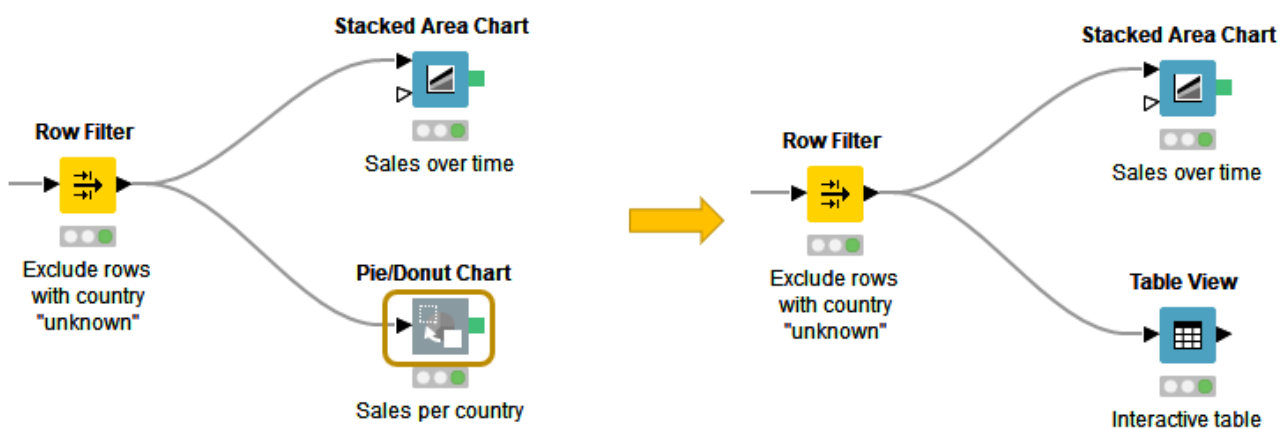


Figure 7. Replacing a node in a workflow

Comments and annotations

You have two options in the workflow editor to document a workflow (as shown in [Figure 8](#)):

- Add a comment to an individual node by double clicking the text field below the node and editing the text
- Add a general comment to the workflow, right click the workflow editor and select *New Workflow Annotation* in the menu. Now a yellow box will appear in the workflow editor.

You can do the following actions on the workflow annotation box.

- To move the workflow annotation box inside the workflow editor, first activate it from the top left corner, and then drag the box.
- To resize the box drag any of its edges.
- To edit the text inside double click the top left corner of the annotation box and type new text in the text field.

- To change the properties of the text and the border double click the top left corner and then right click inside the box. A menu opens showing the available editing options.

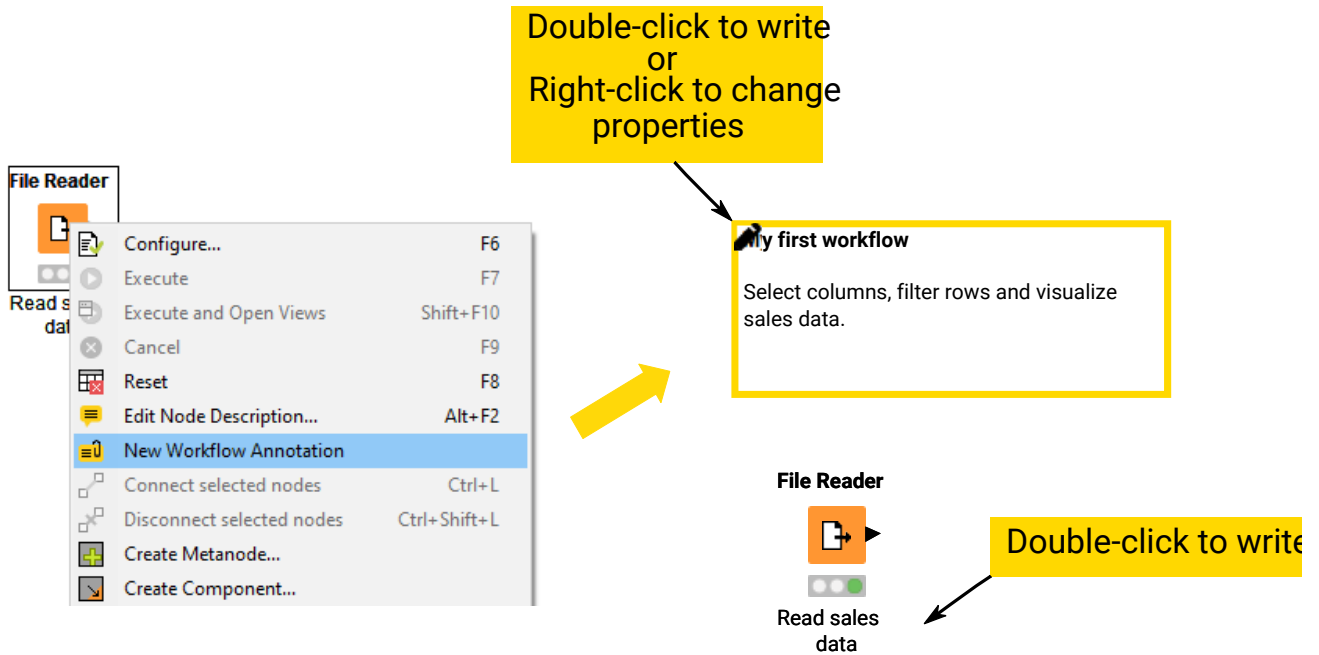


Figure 8. Writing a node comment and creating a workflow annotation



The video [Annotations & Comments](#) gives you a few hints about how to document a workflow.

Workflow editor settings

Change the visual properties of the workflow editor by clicking the "Workflow Editor Settings" button in the toolbar shown in [Figure 9](#).



Figure 9. Changing visual properties of the workflow

In the dialog that opens you can change the size of the grid or remove the grid lines completely. You can also change the connection style from angular to curved, and make the connections thicker or narrower.

The changes will only apply to the currently active workflow editor. To change the default workflow editor settings, go to *File* → *Preferences* → *KNIME* → *KNIME GUI* → *Workflow Editor*.

Keyboard shortcuts

To view a full list of keyboard shortcuts, choose *Help* → *Show Active Keybindings* from the toolbar. Here, it is also possible to modify the bindings, and create personalized shortcuts.

KNIME Explorer

The KNIME Explorer is where you can manage workflows, workflow groups, and server connections. By default only the local workspace, the **EXAMPLES server** and the link to connect to your personal KNIME Hub space are visible in the KNIME Explorer.

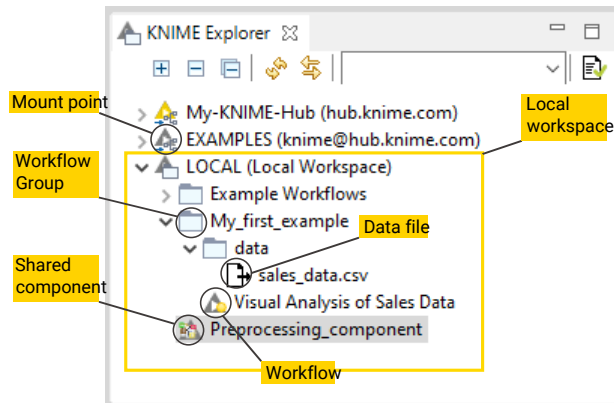


Figure 10. KNIME Explorer

Mount Points

Mount points are workflow repositories that are accessible from KNIME Analytics Platform. They can be displayed as root directories in the KNIME Explorer view.

Each mount point consists of the location of the workflow repository, and a mount ID. For a local workflow repository, the location is the path to the folder, and for a server it is the address of the server. The mount ID is used to reference files and workflows under the mount point.

KNIME Explorer toolbar







At the top of the KNIME Explorer are several icons arranged in a toolbar shown in [Figure 11](#).



Figure 11. KNIME Explorer toolbar

The functions of the icons are explained in [Table 1](#) below:





Table 1. Functions of the KNIME Explorer toolbar icons

	<ul style="list-style-type: none"> • (+) expands the selected workflow group showing its content • (-) collapses the element •  collapses all elements in the KNIME Explorer showing only the mount points
	<p>Refreshes the view, in case it is out of sync with the underlying file system</p>
	<p>Selects the workflow that is open in the workflow editor</p>
	<p>Add text to the field and press "Enter". The KNIME Explorer will only show items that contain the text in their name or are in a workflow group containing the text in its name.</p>
	<p>Opens the explorer preference page, allowing to add/remove/edit mount points</p>

KNIME Explorer content

The types of content that you can see in the KNIME Explorer are described in [Table 2](#).

Table 2. Types of Items in KNIME Explorer

	Workflow	A collection of nodes used to analyze data in KNIME
	Workflow Group	A folder within the KNIME Explorer, which can be used to store workflows, data files, components, and metanodes.
	Data File	Dragging a data file from the KNIME Explorer to the workflow editor automatically creates the correct node to read the file type. Storing data files in the currently active workspace allows for defining file paths relatively to their location in the KNIME Explorer.
	Component/metanode	Components and metanodes contain a pre-configured sub-workflow, which can be integrated in any part of a workflow. Components are nodes that encapsulate and abstract functionality. Metanodes, instead, are used to organize the workflow, collapsing part of it to hide that part of the workflow's functionality.



A more comprehensive overview on components and metanodes is available in the [KNIME Components Guide](#).

Explorer operations

Dragging and dropping elements

In the KNIME Explorer, elements can be moved between the repositories in the same way as in any other file explorer. Besides that, operations can be applied that affect the workflows stored in the KNIME Explorer: create nodes to read different file types and use a component or a metanode within a workflow. These operations are summarized in [Table 3](#).

Table 3. Drag and Drop Operations in the KNIME Explorer

Move	To move an item, simply drag it and drop it to the desired location
Copy	Copying an item is the same process as moving it. Keep the "Ctrl"-key pressed during the drag and drop step. A small plus-sign next to the mouse cursor indicates the copy operation. Additionally, "Ctrl" + "c"/"v" shortcuts can be used to copy and paste elements from one repository to another.
Node creation	Drop a data file into the workflow editor. KNIME will create the appropriate file reading node automatically and preconfigures the node.
Component/metanode usage	A component or a metanode can be saved in the KNIME Explorer for later reuse. To do this, right click any component or metanode and select <i>Component (or Metanode) → Share...</i> . The resulting dialog gives you the possibility to choose a destination and the link type. To use a component or metanode stored in the KNIME Explorer, drag and drop it to the workflow editor.

Context menu

Other useful operations you can do in the KNIME Explorer are available in the context menu. Right click the KNIME Explorer or an item of the view to open the menu shown in [Figure 12](#).

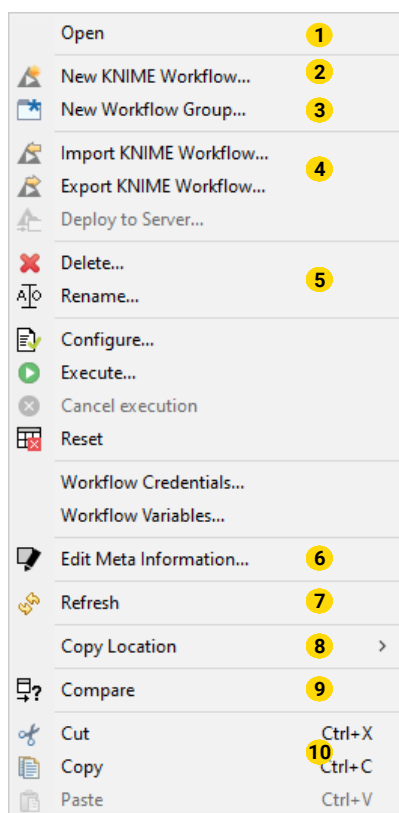


Figure 12. KNIME Explorer Context Menu

The menu items shown in [Figure 12](#) are the operations available for workflows that are stored in your local workspace, without any remote workspace available. The possible operations are listed below:

- 1 Opens the workflow
- 2 **Creates a new, empty workflow**, places it in the selected workflow group and opens it in the workflow editor
- 3 **Creates a new, empty workflow group** and places it in the selected workflow group or directly under the "LOCAL" mount point
- 4 Opens the workflow **import or export wizard**
- 5 Deletes or renames the selected item. If a workflow is currently opened in the workflow editor, or a workflow group contains an open workflow, it is locked and cannot be renamed nor deleted.

- 6 Opens the meta information editor, where it is possible to write a description associated with the selected workflow or workflow group
- 7 Refreshes the workflow (group), in case it is out of sync with the underlying file system
- 8 Expands a menu for **different file path types** to copy the path to the item
- 9 Cuts/copies the selected item

Creating a new workflow

To create an empty workflow, right click anywhere in the local workspace, and select *New KNIME Workflow...* in the menu, or use one of the options explained in [Building Workflows](#). Give the workflow a name, and define the destination of the new workflow.

Click *Finish*, and the new workflow will appear in the selected workflow group in the KNIME Explorer.

To learn how to build a workflow, take a look at the next section [Building Workflows](#), follow the steps in the [Quickstart Guide](#), or check the video [Workflows and Workflow Groups](#).

Building workflows

To create a workflow, you need an empty workflow editor. To create a new empty workflow editor take any of these actions:

- Navigate to *File* → *New...*, and select *New KNIME Workflow*
- Click the leftmost icon in the toolbar
- Right click in the local workspace and select *New KNIME Workflow...*

A workflow is built by dragging nodes from the node repository to the workflow editor and connecting them. To add a node from the node repository or from the workflow coach to the workflow editor, you have two options as shown in [Figure 13](#):

- Drag and drop the node into the workflow editor
- Double click the node

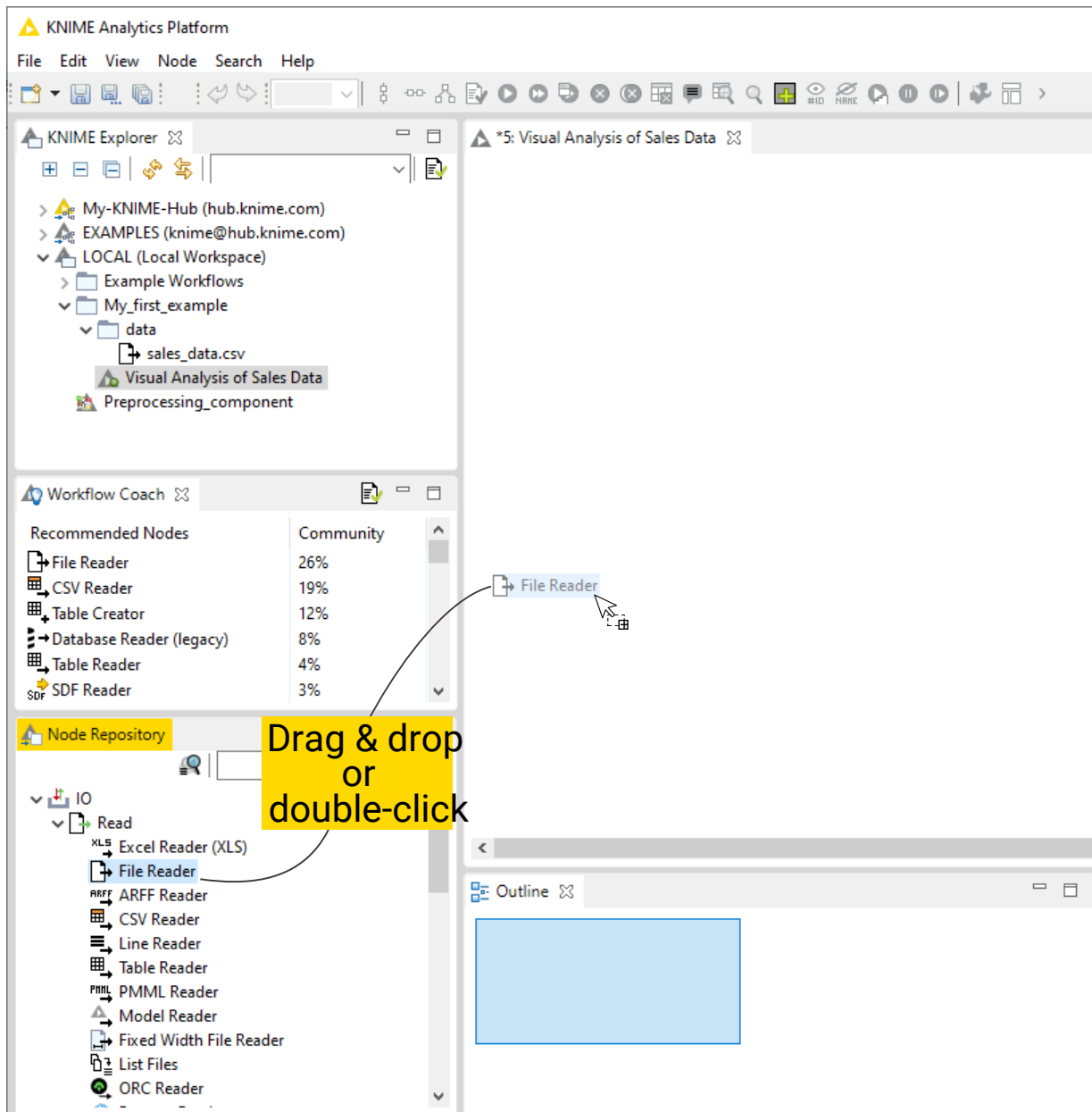


Figure 13. Adding a node into the workflow editor

Once two nodes are added to the workflow editor, they can be connected in any of these three ways:

- Click the output port of the first node and release the mouse at the input port of the second node. Now, the nodes are connected.
- Select a node in the workflow editor, and then double click the next node in the node repository. This double click creates a new node, and connects it to the selected node in the workflow editor.
- Select the nodes to connect in the workflow editor and press "Ctrl + L"

To add a node between two nodes in a workflow, drag the node from the node repository, and release it at its place in the workflow when the connector has turned red, as shown in [Figure 14](#). The red connection means that it is ready to accept the new node. Release the mouse and the node is put in place.

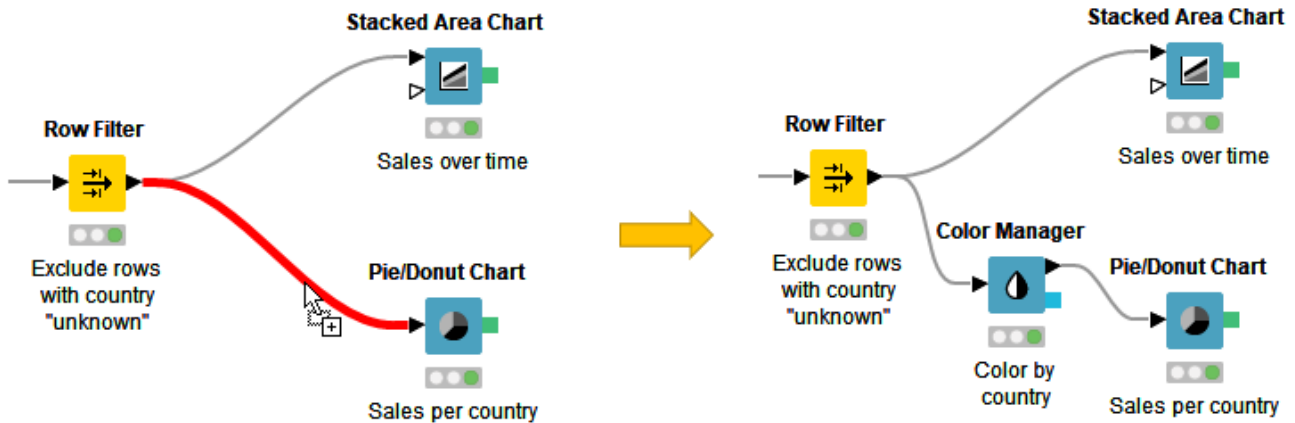


Figure 14. Add a node in the middle of a workflow

Workflow groups

Multiple workflows can be organized into workflow groups. Workflow groups are folders in the KNIME workspace that can include multiple workflows, as well as associated datafiles, shared components and metanodes, and even other workflow groups.

The workflow groups are in the currently active local workspace under the LOCAL mount point in the KNIME Explorer.

You have three ways to create a new, empty workflow group:

- Right click in the local workspace in the KNIME Explorer, and select *New Workflow Group...* in the menu
- Click the arrow next to the leftmost icon in the toolbar and select *New KNIME Workflow Group*
- Navigate to *File* → *New...*, select *New KNIME Workflow Group* in the list, and click *Next*.

In the dialog that opens, give the folder a name, and define where to save the folder in the local workspace. Click *Finish*. Now the new folder will appear in the selected destination in the KNIME Explorer.

Import/export workflows and workflow groups

You have three options to export a workflow or a workflow group:

- Export it as a file
- Save it into your personal KNIME Hub space, either in your public or in your private space
- Or, deploy it to a server (requires a license)



To save workflows into your personal KNIME Hub space you need to sign in. In the KNIME Explorer right click *My-KNIME-Hub (hub.knime.com)* and click *Connect to KNIME Hub*. Please, be aware that when saving a workflow group to your public folder on the KNIME Hub including **data** they become publicly available.

In the same way, you can import a workflow to your local workspace in the following ways:

- Import a file containing a workflow to your local workspace
- Save a workflow that is on a server to your local workspace. For example, you can access the EXAMPLES server (no credentials required) and save any workflow located there to your local workspace.

How to import and export a workflow (or workflow group)

You can import or export a workflow or a workflow group in the following ways:

- Right click anywhere in the local KNIME workspace, and select *Import(Export) KNIME Workflow...*, as shown in [Figure 15](#).

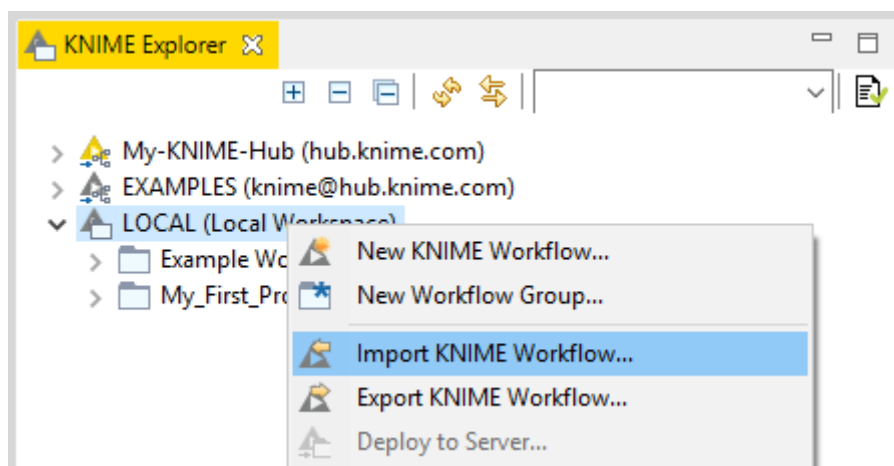


Figure 15. Importing and exporting workflow (groups)

- Go to *File* menu and select *Import (Export) KNIME Workflow...*

The dialog shown in [Figure 16](#) opens.

Importing a workflow

In the upper part of the "Import" dialog, select the items to import, i.e define the file or folder path to import. In the "Destination" field underneath, define the destination folder in the KNIME workspace to import to.

Importing a workflow group will show a list of elements inside the workflow group in the lower part of the dialog. Here you can select single elements to import them.

Exporting a workflow

In the upper part of the "Export" dialog, select the workflow (or group) to export. In the "Destination" field underneath, define the path to the destination folder on the local system, and the name of the file.

In "Options" you can choose to reset the workflow(s) before exporting. After resetting a node, the node status changes from "executed" to "configured", and the output of the node is no longer available. When exporting a workflow in an executed state, the data used in the workflow are exported as well. See the section on [Reset and Logging](#) for more information.

When exporting a workflow group, you can select the elements that you want the exported file to contain.

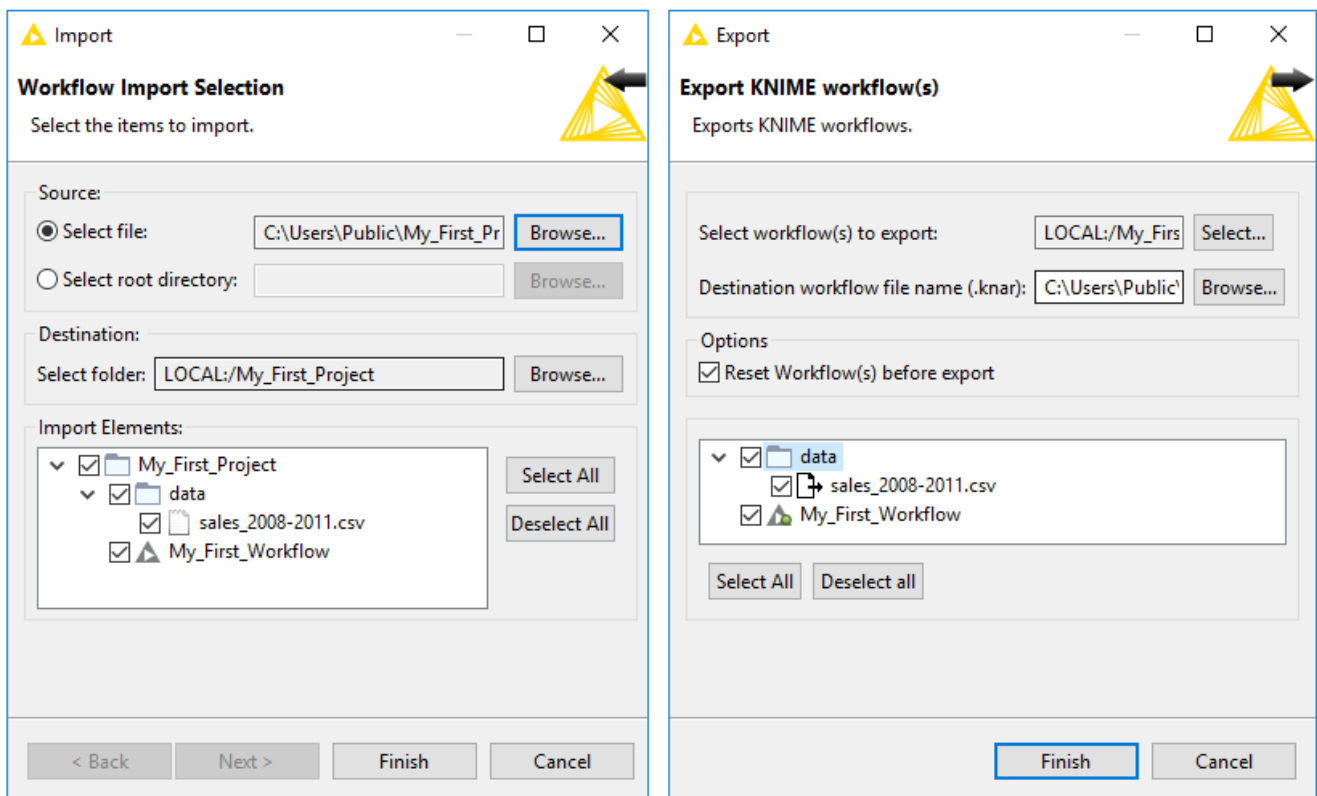


Figure 16. Defining settings for importing and exporting workflow (groups)



Importing and exporting workflows are also introduced in this video: [Import/Export Workflows](#).

- The file extension for a KNIME workflow, is `.knwf` (KNIME workflow file)
- The file extension for a workflow group, is `.knar` (KNIME archive file)

knime:// protocol

knime:// protocol is a protocol specific to KNIME that allows to specify file paths relatively to the KNIME workspace or even the location of the currently executing workflow.

The first element in the file path after knime:// is the base for the path. It is either the workflow itself, the current mount point or a specific mount point like LOCAL in the following example:

```
knime://LOCAL/My_First_Project/data/sales_2008-2011.csv
```



The portable file path options are explained in the subsections below and in this video: [The knime:// Protocol](#).

Absolute URLs

Absolute URLs are defined relative to a specific mount point. The following file path is defined using the absolute path to the file based on the mount point LOCAL:

```
knime://LOCAL/My_First_Project/data/sales_2008-2011.csv
```

The file path would now work on any system where the workflow is saved in the local workspace, and the file path inside the local workspace folder is the same.

Mountpoint-relative URLs

Because of the LOCAL term in the absolute path, accessing the file with the absolute URL is not possible, if the workflow is deployed to a server.

To enable access to a data file both locally and on a server, select the path to the file relative to the currently active mount point.

To do this, change the LOCAL term in the file path to knime.mountpoint as in this file path:

```
knime://knime.mountpoint/My_First_Project/data/sales_2008-2011.csv
```

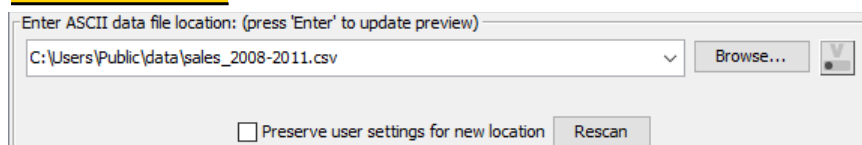
In the mountpoint-relative file path, the knime.mountpoint refers to the uppermost folder level, which can be LOCAL or the mount ID of a server.

Workflow-relative path

The most flexible portable file path is the workflow-relative path. A workflow-relative path defines the file path relative to the currently executing workflow. Using this file path you can access data files in workflows in local workspaces on different systems, or on a server, as long as the folder structure between the workflow and the data file is the same.

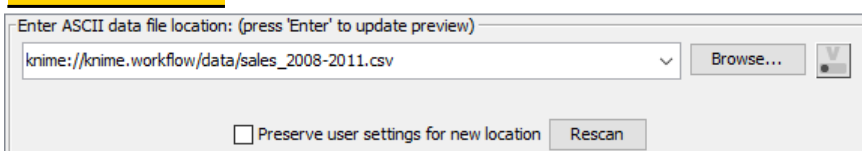
Compared to the absolute path and mountpoint-relative path, the name of the folder containing the workflow does not have to be the same in the different locations. That's because an upper folder level is denoted by `/../` instead of the name of the folder.

Local path



The screenshot shows a dialog box titled "Enter ASCII data file location: (press 'Enter' to update preview)". The text input field contains the local path `C:\Users\Public\data\sales_2008-2011.csv`. To the right of the input field are "Browse..." and "V" buttons. Below the input field, there is a checkbox labeled "Preserve user settings for new location" which is unchecked, and a "Rescan" button.

Workflow-relative path



The screenshot shows the same dialog box as above, but the text input field contains the workflow-relative path `knime://knime.workflow/data/sales_2008-2011.csv`. The "Browse..." and "V" buttons are still present. The "Preserve user settings for new location" checkbox is unchecked, and the "Rescan" button is visible.

Figure 17. Workflow-relative file path

Save workflows with data

You can easily include data into your workflow by using the workflow-relative paths as described above. First, access the workflow in your knime-workspace from your operating system, then manually create a folder called `data`, and place your data inside this folder. In this way you can easily reference your data within nodes using the **workflow-relative path**, which makes sure that your data will remain with your workflow whenever you archive it, export it, or upload it to a KNIME Server or the KNIME Hub.

EXAMPLES server

You can explore the example workflows, which includes also some real-world use cases, on the public EXAMPLES server.

Inspect the workflow groups for different categories by expanding the EXAMPLES mount point in the KNIME Explorer, and then double clicking the text below as shown in **Figure 18**.

You can download an example workflow by drag and drop, or copy and paste of the workflow into the local workspace. Double click the downloaded copy of the example workflow to open and edit it like any other workflow.

Alternatively, double click the example workflow directly on the EXAMPLES server to open it in the workflow editor. Save it to the local workspace via "File" and then "Save As...".

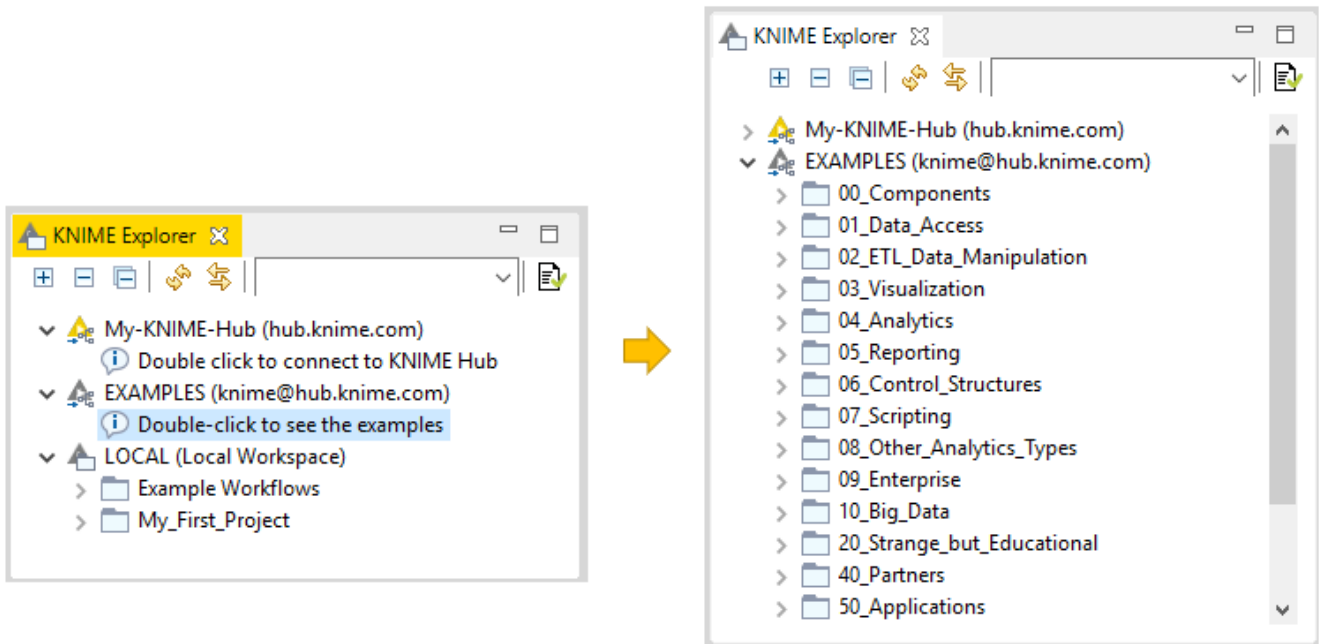


Figure 18. Example workflow available on the EXAMPLES server



The video [The EXAMPLES server](#) provides a more detailed introduction to the EXAMPLES server.

Workflow Coach

The workflow coach shown in [Figure 19](#) provides node recommendations. If a node is selected in the workflow editor, the workflow coach shows the most popular nodes to follow the selected node. Otherwise, the recommendations represent the most popular nodes to start a workflow.

The recommendations are based on KNIME community usage statistics about workflows built in KNIME Analytics Platform. Nodes can be added from the workflow coach to the workflow editor in the same way as from the node repository, by drag and drop, or by a double click.

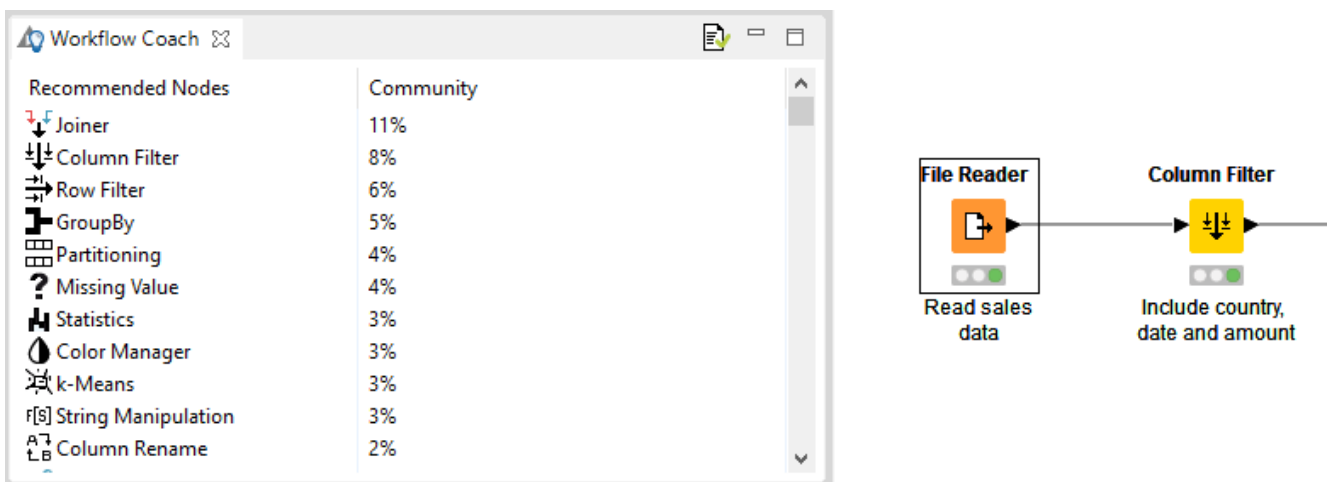


Figure 19. Workflow Coach

Note: start or stop sending anonymous usage data any time by checking the option *Yes, help improve KNIME.* in the "KNIME" dialog in [Preferences](#).

Customizing node recommendations

Customize the node recommendations in the "Workflow Coach" dialog, under *File* → *Preferences* → *KNIME* → *Workflow Coach*. You have the following three options:

- Add node recommendations based on workflows in the currently active local workspace by enabling the *Workspace Node Recommendations* option in the "Workspace Recommendations" dialog
- Add node recommendations based on the workflows on a server by selecting the KNIME Server in the "Server Recommendations" dialog
- Disable the node recommendations by the community by unchecking the *Node Recommendations by the Community* option in the "Workflow Coach" dialog



The video [Workflow Coach: The Wisdom of the KNIME Crowd](#) provides a more detailed introduction to node recommendations.

Node repository

Currently installed nodes are available in the node repository where they are organized under different categories. You can add a node from the node repository into the workflow editor by drag and drop, or by a double click, as explained in the section [Building Workflows](#).

Search for a node by expanding the categories or by typing a search term in the search field on top of the node repository, as shown in [Figure 20](#). The default search mode is *crisp search*. Using this search mode, the interface returns all the nodes that either have the search term in their names, or are in a subcategory whose name includes the search term.

Switch the search mode to *fuzzy search* by clicking the icon next to the search field. In this search mode the interface returns all the nodes that are related to the search term.

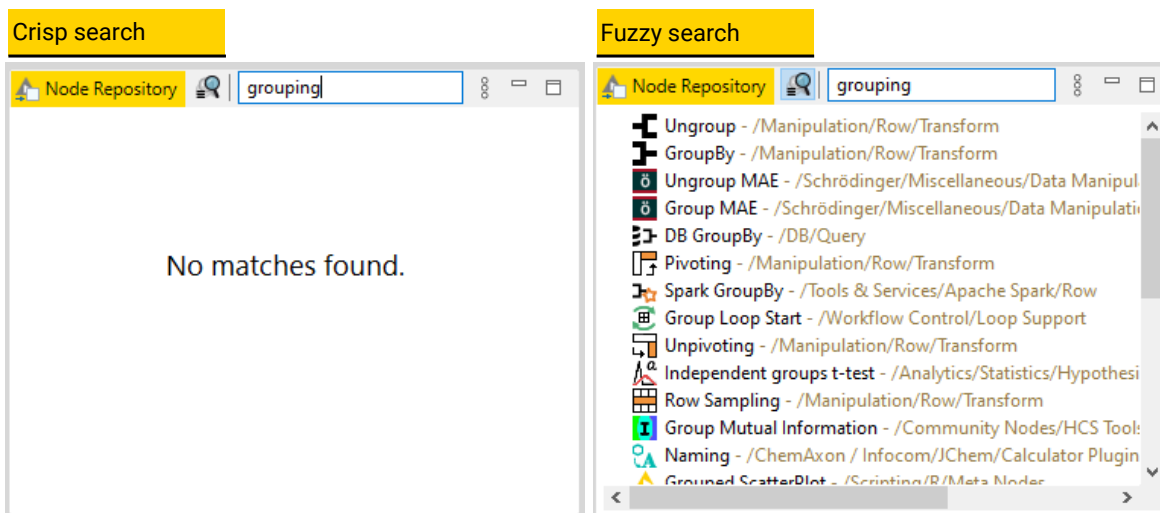


Figure 20. Node repository with two search modes



An introduction to the node repository is also available in the video [Node Repository](#).

KNIME Hub Search

The search bar on the right of the KNIME Workbench shown in [Figure 2](#) provides a convenient way to directly search the [KNIME Hub](#) from within KNIME Analytics Platform. After entering a search term or sentence, press "Enter". This will open a browser window and display the results of the search as shown in [Figure 21](#).

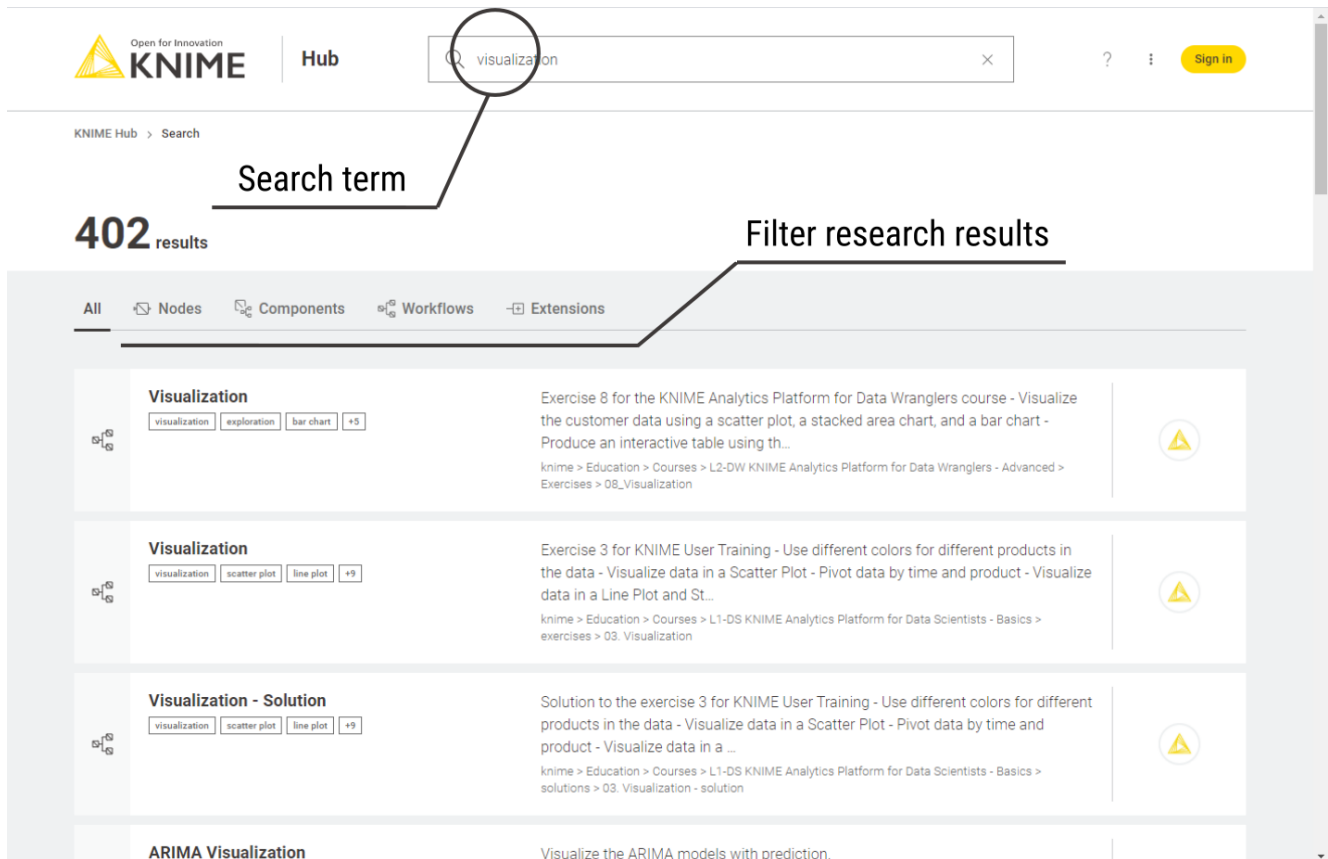


Figure 21. Inspecting the KNIME Hub Search results on a browser

From the KNIME Hub page, you can:

- Inspect, open, and download workflows
- Add nodes and components into the workflow by drag and drop
- Install extensions by drag and drop

Description

The description panel on the right of the KNIME Workbench shown in [Figure 2](#) provides a description of the currently active workflow, or a node selected in the node repository or workflow editor. For a workflow, the first part is a general description, followed by tags and links to other resources related to the workflow. For a node, the first part is a general description, followed by the available setting options, and finally a list of input and output ports.

Node Monitor

The Node Monitor tab is located on the same panel of the [console](#) tab on the bottom part of the KNIME Workbench shown in [Figure 22](#). It is especially useful to inspect intermediate output tables in the workflow.



The Node Monitor tab is shown by default since KNIME Analytics Platform version 4.2. For KNIME Analytics Platform version <4.1, or in case you closed the Node Monitor tab and you want to restore it, go to *View* in the toolbar and select *Node Monitor* from the menu.

ID	country	date	amount
Row0	unknown	2008-12-12	3
Row1	China	2009-04-10	160
Row2	China	2009-04-10	160
Row3	China	2009-05-10	160
Row4	USA	2009-05-20	1600
Row5	Brazil	2009-06-08	1200
Row6	USA	2009-07-04	70
Row7	USA	2009-07-14	70
Row8	USA	2009-08-20	1600
Row9	Germanv	2009-11-02	600

Figure 22. Node Monitor

Here you can choose to show the flow variables or a preview of the output data at any port of a selected node in the active workflow. To choose what kind of output you want to visualize, click the three vertical dots on the right-upper side of the node monitor tab, and select the You can also pin the monitor view to a specific node independently of the selection in the workflow editor, in order, for example, to follow the evolution of data flow and help with debugging. To do this select the node you want to pin, and click the green pin symbol on the right-upper side of the node monitor tab.

Outline

In the outline, on the bottom part of the KNIME Workbench shown in [Figure 23](#), you find an overview of the currently active workflow. If the whole workflow does not fit in the workflow editor, you can change the active area by scrolling the blue, transparent rectangle.

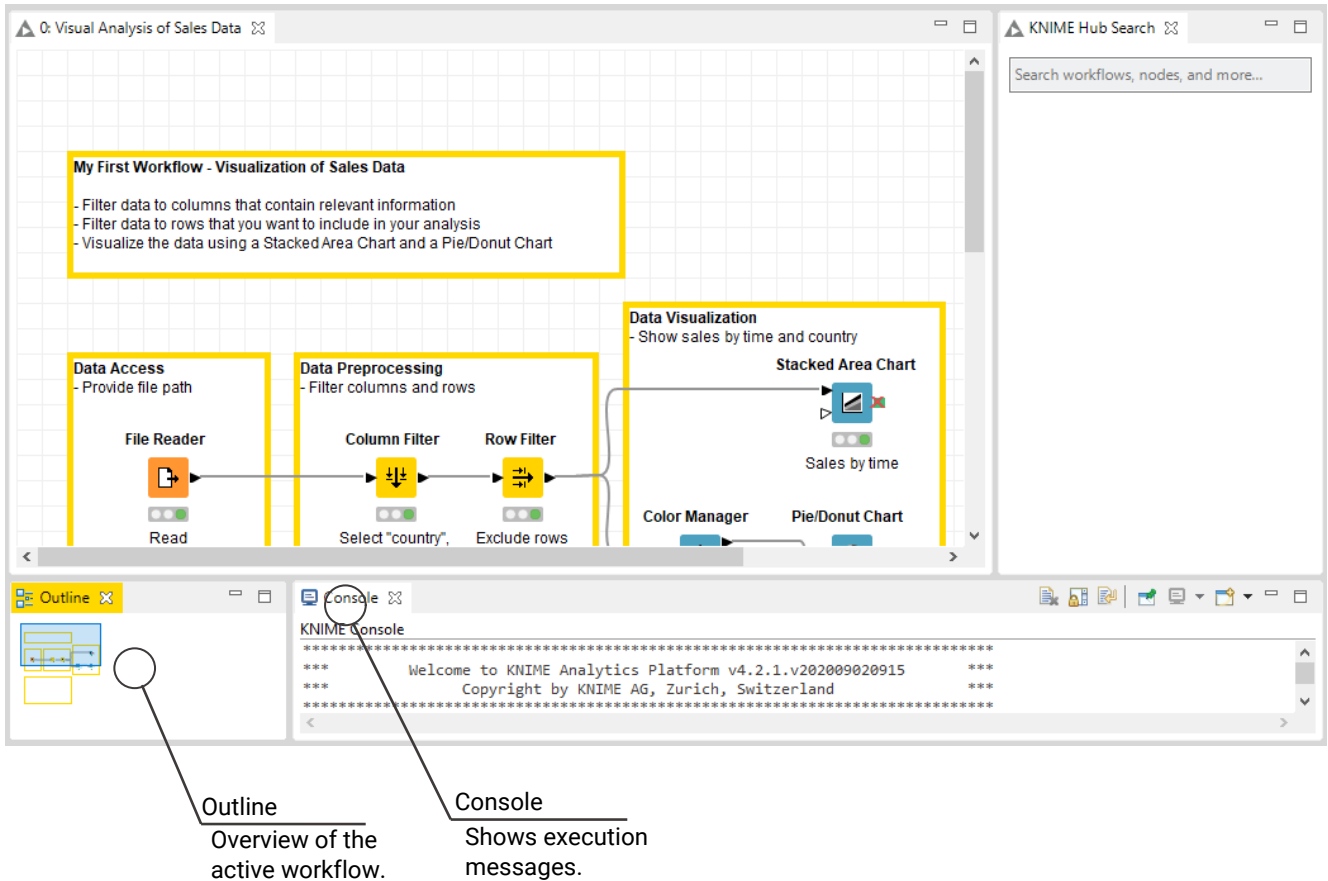


Figure 23. Outline and console

Console

The console tab on the bottom part of the KNIME Workbench shown in [Figure 23](#) shows all warning and error messages related to the workflow execution. To debug and log information messages to be reported in the console, change the console log level in *File* → *Preferences* → *KNIME* → *KNIME GUI*.

Customizing the KNIME Workbench

Reset and logging

When a node is reset, the node status changes from "executed" to "configured" and the output of the node is not available any more. When saving a workflow in an executed state, the data used in the workflow are saved as well. That is, the larger the dataset, the larger the file size. Therefore, resetting workflows before saving them is recommended in case the dataset can be accessed without any restrictions.

A reset workflow only saves the node configurations, and not any results. However, resetting a node does not undo the operation executed before. All operations done during creation, configuration, and execution of a workflow are reported in the `knime.log` file.

To inspect the `knime.log` file you go to *View* → *Open KNIME log*. The log file opens in the workflow editor. The `knime.log` file has a limited size, and after reaching it the rows will be overwritten from the top.

The `knime.log` file is also located in the `knime`-folder inside the `.metadata`-folder, in the KNIME workspace folder defined when launching KNIME Analytics Platform.

Show heap status

The heap status panel shows the memory usage during the execution of a workflow, and helps to monitor memory usage for the project. To add the heap status panel to the workbench, go to *File* → *Preferences*. In the dialog that opens, click *General*, select *Show heap status*, and click *Apply and Close*.

A heap status bar showing the memory usage appears in the bottom right part of the status bar, directly below the console panel. Next to the heap status bar is the "Run Garbage Collector" button. Click it to manually allocate and free up memory.

Configuring KNIME Analytics Platform

Preferences

In *Preferences* you can adjust the default settings of KNIME Analytics Platform. Go to *File* → *Preferences*, and a list of subcategories is displayed in the dialog that opens. Each category contains a separate dialog for specific settings like database drivers, available update sites, and appearance.

KNIME

Selecting *KNIME* in the list of subcategories, allows you to define the log file log level. By default it is set to *DEBUG*. This log level helps developers to find reasons for any unexpected behavior.

Directly below, you can define the maximum number of threads for all nodes. Separate branches of the workflow are distributed to several threads to optimize the overall execution time. By default the number of threads is set to twice the number of CPUs on the running machine.

In the same dialog, you can also define the folder for temporary files.

Check the last option *Yes, help improve KNIME.* to agree to sending us anonymous usage data. This agreement activates the node recommendations by community in the [Workflow Coach](#).

KNIME GUI

The *KNIME* category, contains a subcategory *KNIME GUI*. In this dialog, you can define the console view log level. By default it is set to "WARN", because more detailed information is only useful for diagnosis purposes.

Further below, you can select which confirmation dialogs are shown when using KNIME Analytics Platform. Choose from the following:

- Confirmation after resetting a node
- Deleting a node or connection
- Replacing a connection
- Saving and executing workflow

- Loading workflows created with a nightly build

In the same dialog, you can define what happens if an operation requires executing the previous nodes in the workflow. You have these three options:

- Execute the nodes automatically
- Always reject the node execution
- Show a dialog to execute or not

The following options allow you to define whether workflows should be saved automatically and after what time interval, also whether linked components and metanodes should be automatically updated. You can also define visual properties such as the border width of workflow annotations.

Master key

Any credentials in use in a workflow can be encrypted using a master key. Once, for example, you enter credentials for different database connections in a workflow, you do not need to save them together with the workflow, nor you need to enter them every time the workflow is opened. Instead, you just need to provide the master key.

Setting up knime.ini

When installing KNIME Analytics Platform, configuration options are set to their defaults. The configuration options, i.e. options used by KNIME Analytics Platform, range from memory settings to system properties required by some extensions.

You can change the default settings in the `knime.ini` file. The `knime.ini` file is located in the installation folder of KNIME Analytics Platform.



To locate the `knime.ini` file on MacOS, open Finder and navigate to the installed Applications.

Next, right click the KNIME application, select *Show Package Contents* in the menu, and navigate to *Contents*, and open *Eclipse*.

Edit the `knime.ini` file with any plaintext editor, such as Notepad (Windows), TextEdit (MacOS) or gedit (Linux).

The entry `-Xmx1024m` in the `knime.ini` file specifies how much memory KNIME Analytics Platform is allowed to use. The setting for this value will depend on how much memory is available in the running machine. We recommend setting it to approximately one half of the

available memory, but this value can be modified and personalized. For example, if the computer has 16GB of memory, the entry might be set to `-Xmx8G`.

Besides the memory available, you can define many other settings in the `knime.ini` file. Find an overview of some of the most common settings in [Table 4](#) or in this [complete list](#) of the configuration options.

Table 4. Common configuration settings in `knime.ini` file

Setting	Explanation
<code>-Xmx</code> <ul style="list-style-type: none"> • default value: 1024m • example: <code>-Xmx16G</code> 	Sets the maximum amount of memory available for KNIME Analytics Platform.
<code>-Dknime.compress.io</code> <ul style="list-style-type: none"> • default value: SNAPPY • possible values: [SNAPPY GZIP NONE] • example: <code>-Dknime.compress.io=SNAPPY</code> 	Determines which compression algorithm (if any) to use when writing temporary tables to disk.
<code>-Dorg.knime.container.cellsinmemory</code> <ul style="list-style-type: none"> • default value: 5,000 • possible values: any value between 0 and 2,147,483,647 • example: <code>-Dorg.knime.container.cellsinmemory=100,000</code> 	This setting defines the size of a "small table". Small tables are attempted to be kept in memory, independent of the Table Caching strategy. By increasing the size of a small table, the number of swaps to the disk can be limited, which comes at the cost of reducing memory space available for other operations.
<code>-Dknime.table.cache</code> <ul style="list-style-type: none"> • default value: LRU • possible values: [LRU SMALL] • example: <code>-Dknime.table.cache=SMALL</code> 	Determines whether to attempt to cache large tables (i.e., tables that are not considered to be "small"; see setting <code>-Dorg.knime.container.cellsinmemory</code>) in memory. If set to LRU, large tables are evicted from memory in least-recently used (LRU) order or when memory becomes scarce. If set to SMALL, large tables are always flushed to disk.

Setting	Explanation
<p data-bbox="150 286 454 322"><code>-Dknime.url.timeout</code></p> <ul data-bbox="204 362 737 465" style="list-style-type: none"><li data-bbox="204 362 560 398">• default value: 1,000 ms<li data-bbox="204 430 737 465">• example: <code>-Dknime.url.timeout=100</code>	<p data-bbox="817 286 1444 555">When trying to connect or read data from an URL, this value defines a timeout for the request. Increase the value if a reader node fails. A too high timeout value may lead to slow websites blocking dialogs in KNIME Analytics Platform.</p>

KNIME tables

Data table

Very common input and output ports of **nodes** are data input ports and data output ports, which correspond to the black triangles in **Figure 24**.

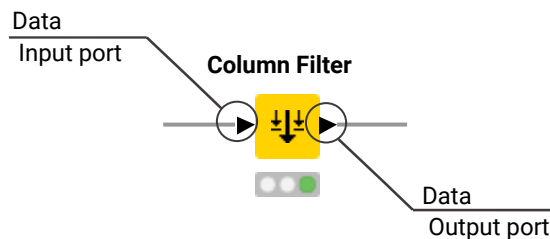


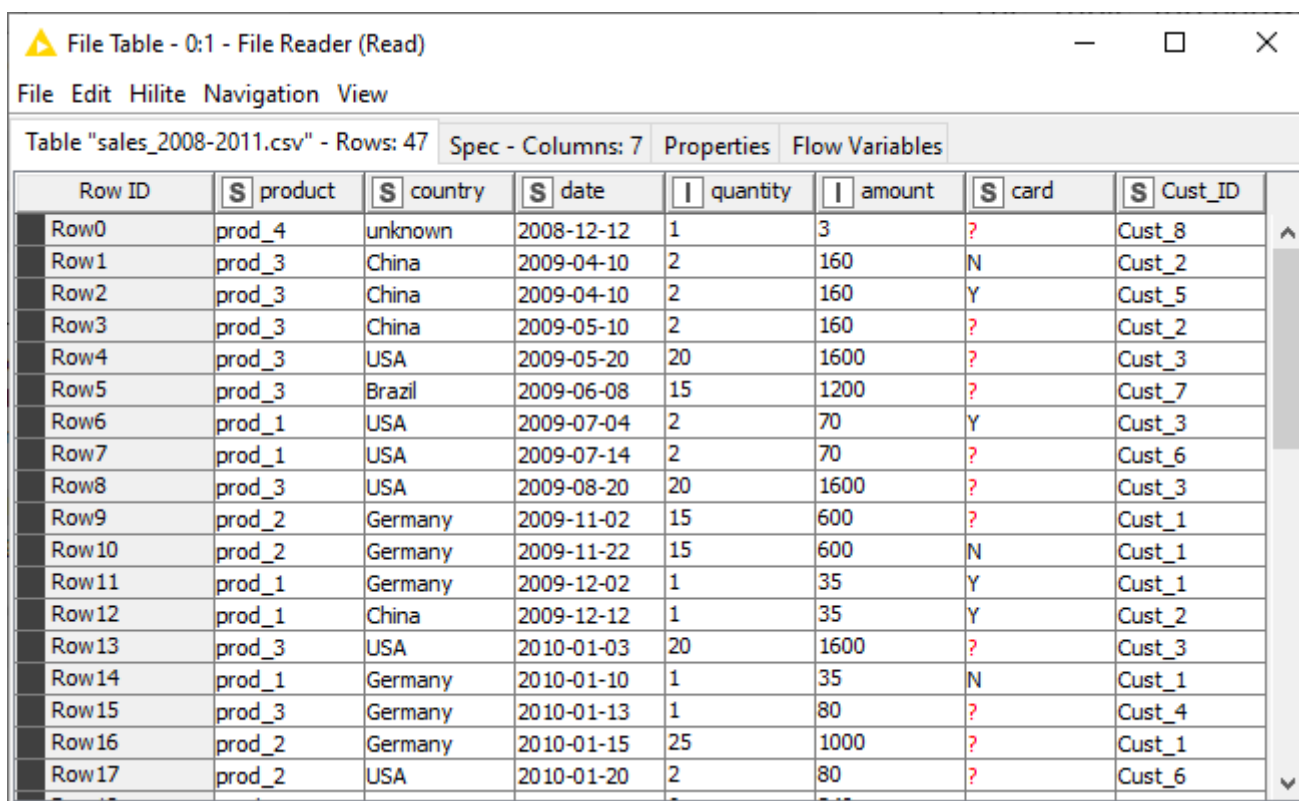
Figure 24. Data input and output port

A data table is organized by columns and rows, and it contains a number of equal-length rows. Elements in each column must have the same data type.

The data table shown in **Figure 25** is produced by the File Reader node, which is one of the many nodes with a black triangle output port for data output. To open the table, right click the node and select the last item *File Table* in the menu. The output table has unique row IDs and column headers. The row IDs are automatically created by the reader node, but they can also be defined manually. The row IDs and the column headers can therefore be used to identify each data cell in the table. Missing values in the data are shown by a question mark.

Besides the data table, the node output contains the following tabs:

1. The "Table" tab shows the contents of the table
2. The "Spec" tab shows the meta information of the table, including the column name, column type, and optional properties like the domain of the values in the column
3. The "Properties" tab, shows metadata related to some columns, for example the width of the histogram in the "Histogram" column produced by the Statistics node
4. The "Flow Variables" tab shows the available flow variables in the node output and their current values.



Row ID	S product	S country	S date	I quantity	I amount	S card	S Cust_ID
Row0	prod_4	unknown	2008-12-12	1	3	?	Cust_8
Row1	prod_3	China	2009-04-10	2	160	N	Cust_2
Row2	prod_3	China	2009-04-10	2	160	Y	Cust_5
Row3	prod_3	China	2009-05-10	2	160	?	Cust_2
Row4	prod_3	USA	2009-05-20	20	1600	?	Cust_3
Row5	prod_3	Brazil	2009-06-08	15	1200	?	Cust_7
Row6	prod_1	USA	2009-07-04	2	70	Y	Cust_3
Row7	prod_1	USA	2009-07-14	2	70	?	Cust_6
Row8	prod_3	USA	2009-08-20	20	1600	?	Cust_3
Row9	prod_2	Germany	2009-11-02	15	600	?	Cust_1
Row10	prod_2	Germany	2009-11-22	15	600	N	Cust_1
Row11	prod_1	Germany	2009-12-02	1	35	Y	Cust_1
Row12	prod_1	China	2009-12-12	1	35	Y	Cust_2
Row13	prod_3	USA	2010-01-03	20	1600	?	Cust_3
Row14	prod_1	Germany	2010-01-10	1	35	N	Cust_1
Row15	prod_3	Germany	2010-01-13	1	80	?	Cust_4
Row16	prod_2	Germany	2010-01-15	25	1000	?	Cust_1
Row17	prod_2	USA	2010-01-20	2	80	?	Cust_6

Figure 25. Data output in KNIME Analytics Platform



In the video [Data Table Structure](#) we introduce the data organization and data representation in KNIME Analytics Platform in more detail.

Column types

The basic data types in KNIME Analytics Platform are Integer, Double, and String, along with other supported data types such as Long, Boolean value, JSON, URI, Document, Date&Time, Bit vector, Image, and Blob. KNIME Analytics Platform also supports customized data types, for example, a representation of a molecule.

Click the "Spec" tab in an output table, to see the data types of the columns in the data table, as well as the domain of the values in the columns, as shown in [Figure 26](#). For numerical values, only the range of the values in the data is shown. For string values, the different values appearing in the data are shown.

Colu...	Column Type	Column Index	Lower Bound	Upper Bound	Value 0	Value 1	Value 2	Value 3	Value 4	Value 5
product	String	0				?	?	prod_4	prod_3	prod_1	prod_2	?	?
country	String	1				?	?	unknown	China	USA	Brazil	Germany	?
date	String	2				?	?	2008-12-12	2009-04-10	2009-05-10	2009-05-20	2009-06-08	2009-0
quantity	Number (int...	3				1	25	?	?	?	?	?	?
amount	Number (int...	4				3	1600	?	?	?	?	?	?
card	String	5				?	?	N	Y	?	?	?	?
Cust_ID	String	6				?	?	Cust_8	Cust_2	Cust_5	Cust_3	Cust_7	Cust_6

Figure 26. Data types and data domain in "Spec" tab

The reader nodes in KNIME Analytics Platform assign a data type to each column based on their interpretation of the content. If the correct data type of a column is not recognized by the reader node, the data type can be corrected afterwards. There are nodes available to convert data types. For example: String to Number, Number to String, Double to Int, String to Date&Time, String to JSON, and String to URI.

Many of the special data types are recognized as `String` by the reader nodes. To convert these `String` columns to their correct data types, use the `Column Type Auto Cast` node.

When you use the File Reader node to read a file you can convert the column types directly via the node configuration dialog. To do this double click a column header in the preview and change the column type in the dialog that opens, as shown in [Figure 27](#).

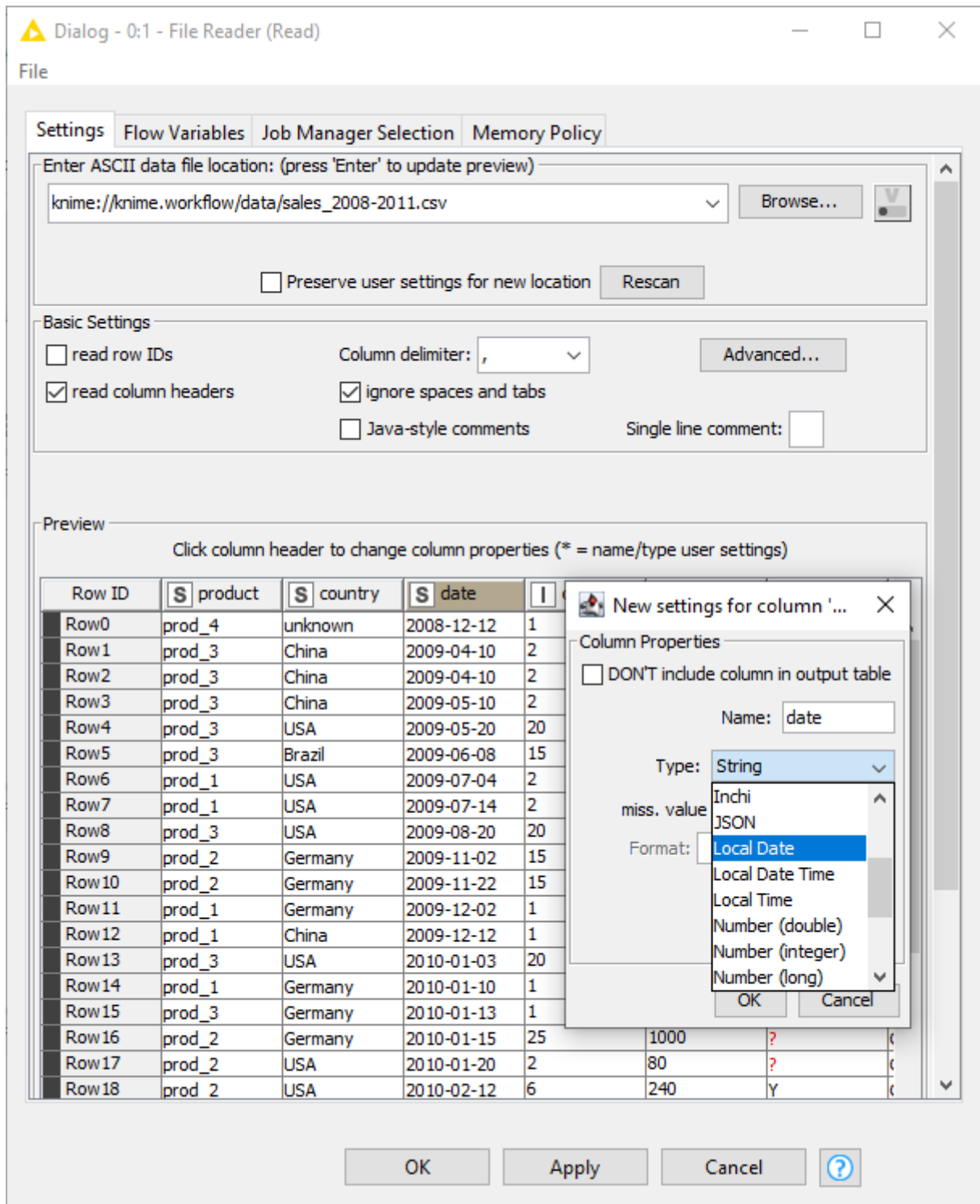
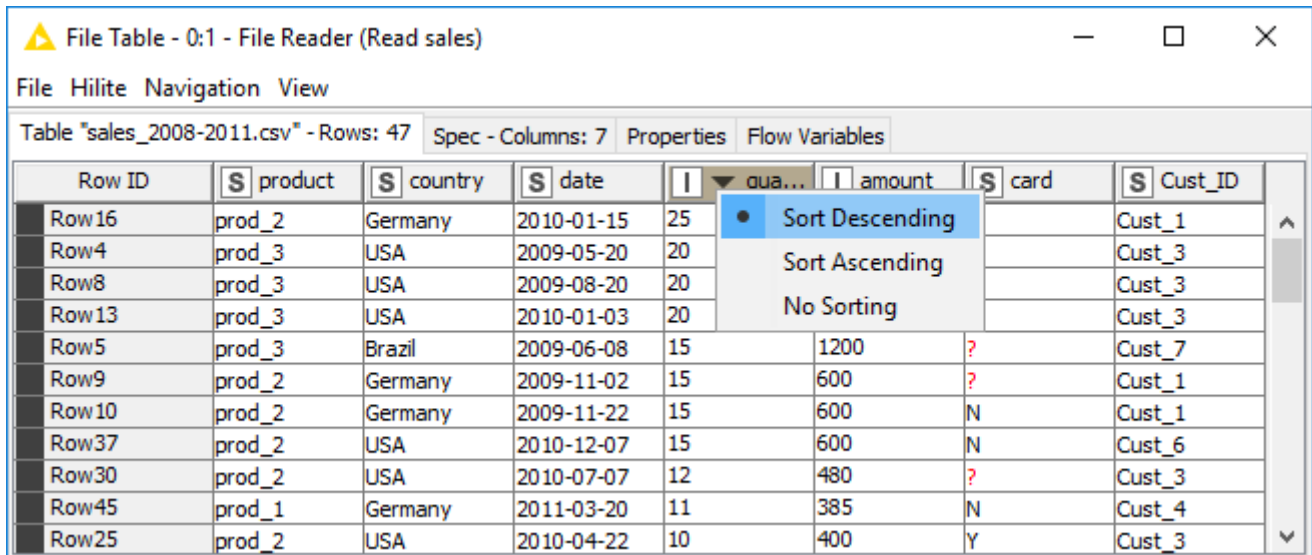


Figure 27. Change column type in File Reader node

Sorting

Rows in the table view output can be sorted by values in one column by clicking the column header and selecting *Sort Descending* or *Sort Ascending* as shown in [Figure 28](#). Note that this sorting only affects the current output view and has no effect on the node output.



The screenshot shows a window titled "File Table - 0:1 - File Reader (Read sales)". The table view displays data from "Table 'sales_2008-2011.csv' - Rows: 47". The columns are: Row ID, product, country, date, quantity, amount, card, and Cust_ID. A context menu is open over the 'quantity' column header, showing three options: "Sort Descending" (selected), "Sort Ascending", and "No Sorting".

Row ID	product	country	date	quantity	amount	card	Cust_ID
Row16	prod_2	Germany	2010-01-15	25			Cust_1
Row4	prod_3	USA	2009-05-20	20			Cust_3
Row8	prod_3	USA	2009-08-20	20			Cust_3
Row13	prod_3	USA	2010-01-03	20			Cust_3
Row5	prod_3	Brazil	2009-06-08	15	1200	?	Cust_7
Row9	prod_2	Germany	2009-11-02	15	600	?	Cust_1
Row10	prod_2	Germany	2009-11-22	15	600	N	Cust_1
Row37	prod_2	USA	2010-12-07	15	600	N	Cust_6
Row30	prod_2	USA	2010-07-07	12	480	?	Cust_3
Row45	prod_1	Germany	2011-03-20	11	385	N	Cust_4
Row25	prod_2	USA	2010-04-22	10	400	Y	Cust_3

Figure 28. Sorting data in table view

To sort rows in an output table permanently, use the Sorter node. Use the Column Resorter node to reorder columns.

Column rendering

In a table view output, you can also change the way in which numeric values are displayed in a data table. For example, it is possible to display numeric values as percentages, with full precision, or replace digits by a color scale or bars. To see these and other rendering options for a column, right click the column header, and select *Available Renderers* as shown in [Figure 29](#). Note that these changes are temporary and have no effect on the node output.

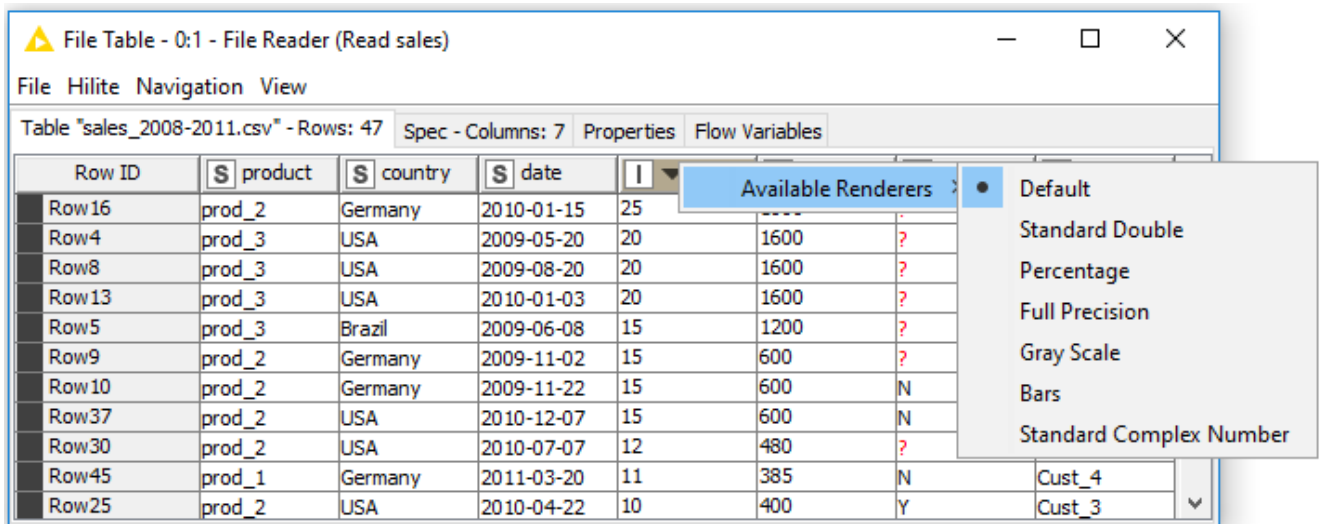


Figure 29. Rendering data in table view

Table storage

When executed, many KNIME nodes generate and provide access to tabular data at their output ports. These tables might be small or large and, therefore, might fit into the main memory of the executing machine or not. Several options are available for configuring which tables to hold in memory as well as when and how to write tables to disk. These options are outlined in this section.

In-memory caching

KNIME Analytics Platform differentiates between small and large tables. Tables are considered to be small (large) when they are composed of up to (more than) 5000 cells. This threshold of 5000 cells can be adjusted via the `-Dorg.knime.container.cellsinmemory` parameter in the `knime.ini` file. KNIME Analytics Platform always attempts to hold small tables in memory, flushing them to disk only when memory becomes scarce.

In addition, KNIME Analytics Platform attempts to keep recently used large tables in memory while sufficient memory is available. However, it writes these tables asynchronously to disk in the background, such that they can be dropped from memory when they have not been accessed for some time or when memory becomes scarce. You can **configure** the memory consumption of a specific node to never attempt to hold its tables in memory and, instead, write them to disk on execution. This is helpful if you know that a node will generate a table that cannot be held in memory or if you want to reduce the memory footprint of a node.

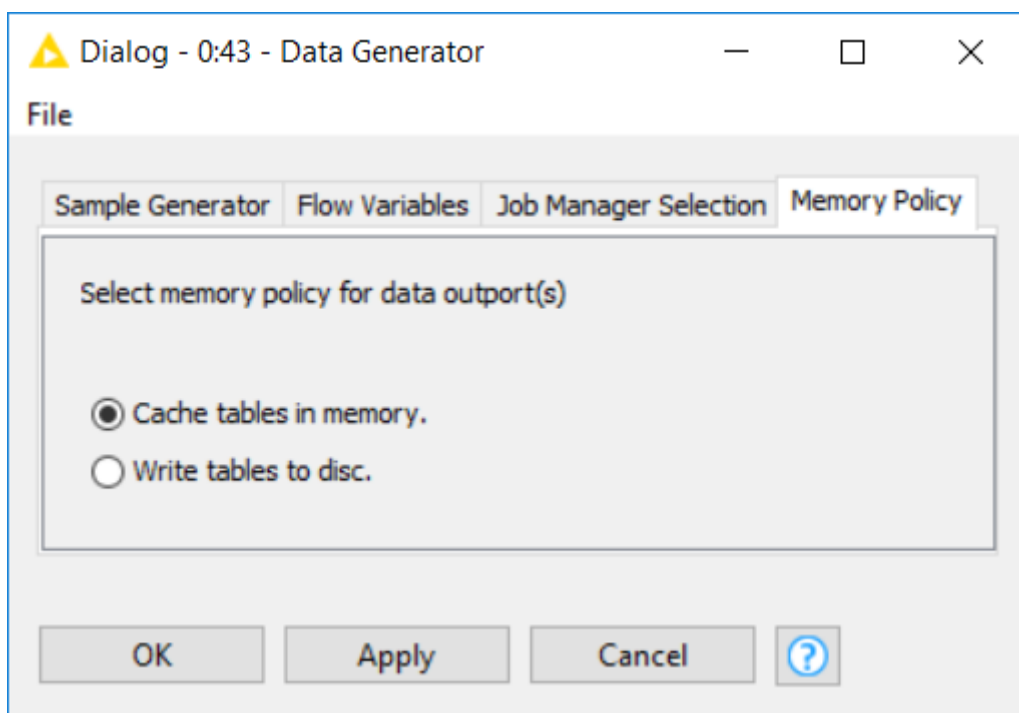


Figure 30. Configuring a node's memory policy

Alternatively, by putting the line `-Dknime.table.cache=SMALL` into the `knime.ini` file, KNIME Analytics Platform can be globally configured to use a less memory-consuming, albeit much slower caching strategy. This strategy only ever keeps small tables in memory.

Disk storage

KNIME Analytics Platform compresses tables written to disk to reduce the amount of occupied disk space. By default, KNIME Analytics Platform uses the **Snappy compression** algorithm to compress its tables. However, you can configure KNIME Analytics Platform to use GZIP compression or no compression scheme at all via the `-Dknime.compress.io` parameter in the `knime.ini` file.

KNIME Analytics Platform also supports columnar storage of tables using the **Apache Parquet** format. The columnar table storage format provides much faster access to individual sections of a table and can therefore speed up execution of certain nodes. To install the columnar data storage, in the KNIME Analytics Platform, go to *File* → *Install KNIME Extensions...* and search for *KNIME Column Storage (based on Apache Parquet)*. To enable the columnar table storage, navigate to *File* → *Preferences* → *KNIME* → *Data Storage*.

KNIME AG
Talacker 50
8001 Zurich, Switzerland
www.knime.com
info@knime.com