KNIME

nodes blending

Alessandra Santi

Linux Day a Pisa 2022

22 ottobre 2022

Open for Innovation
KNIME

KNIME (pronunciato /naɪm/) è una piattaforma open source con licenza GPLv3 di analisi dati, reportistica e integrazione. Ha anche al suo interno componenti di machine learning e data mining.

fonte: https://it.wikipedia.org/wiki/KNIME

Open for Innovation
KNIME

L’interfaccia grafica di KNIME permette di assemblare i nodi per la preprocessazione dei dati

ETL:

- **E**xtraction
- **T**ransformation
- **L**oading

per la modellazione, l'analisi dei dati e la visualizzazione.

fonte: https://it.wikipedia.org/wiki/KNIME

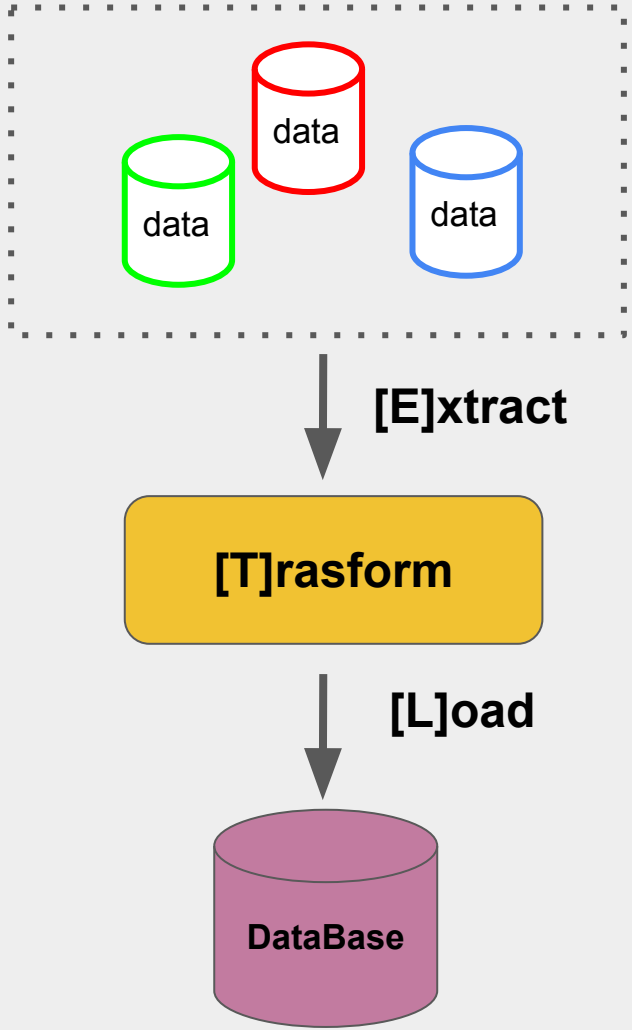
Che cos’è l’ ETL ?

Open for Innovation
KNIME

Che cos’è l’ ETL?

- **[E]**xtract(ion) = Estrarre
- **[T]**ransform(ation) = Trasformare
- **[L]**oad(ing) = Caricare

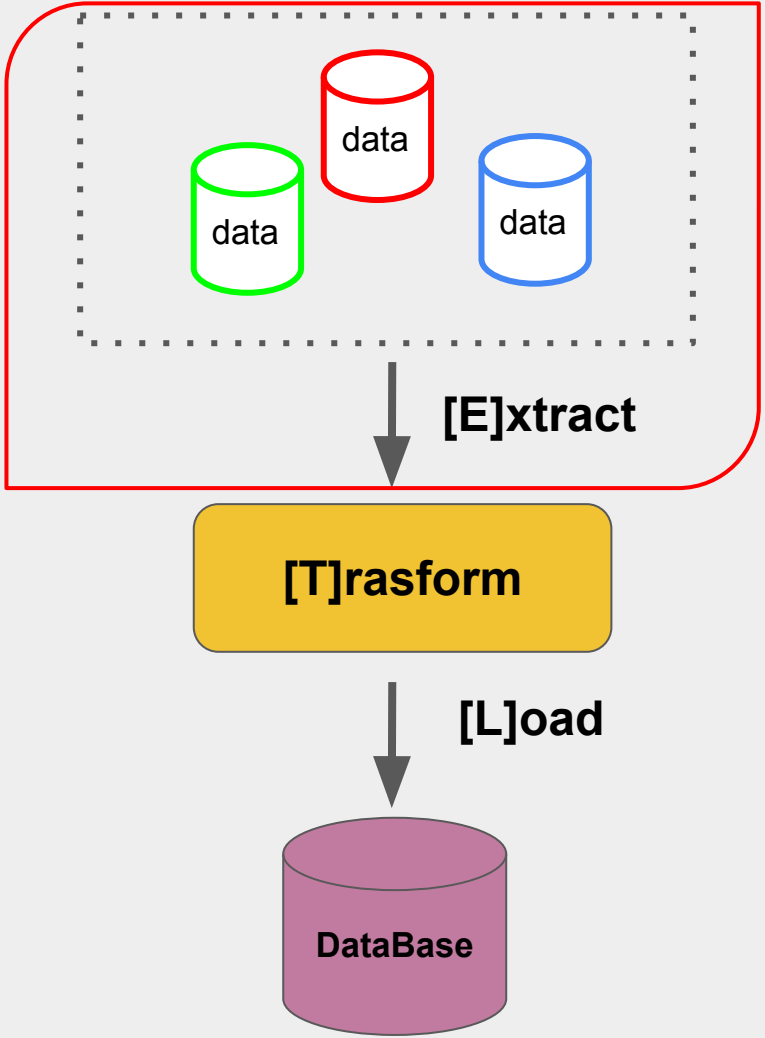
Processo che recupera i dati da più fonti **[E]**, li trasforma in formato standard **[T]** e li ricarica su un’altra fonte database **[L]**



Open for Innovation
KNIME

ETL →**[E]**xtract(ion) = Estrarre

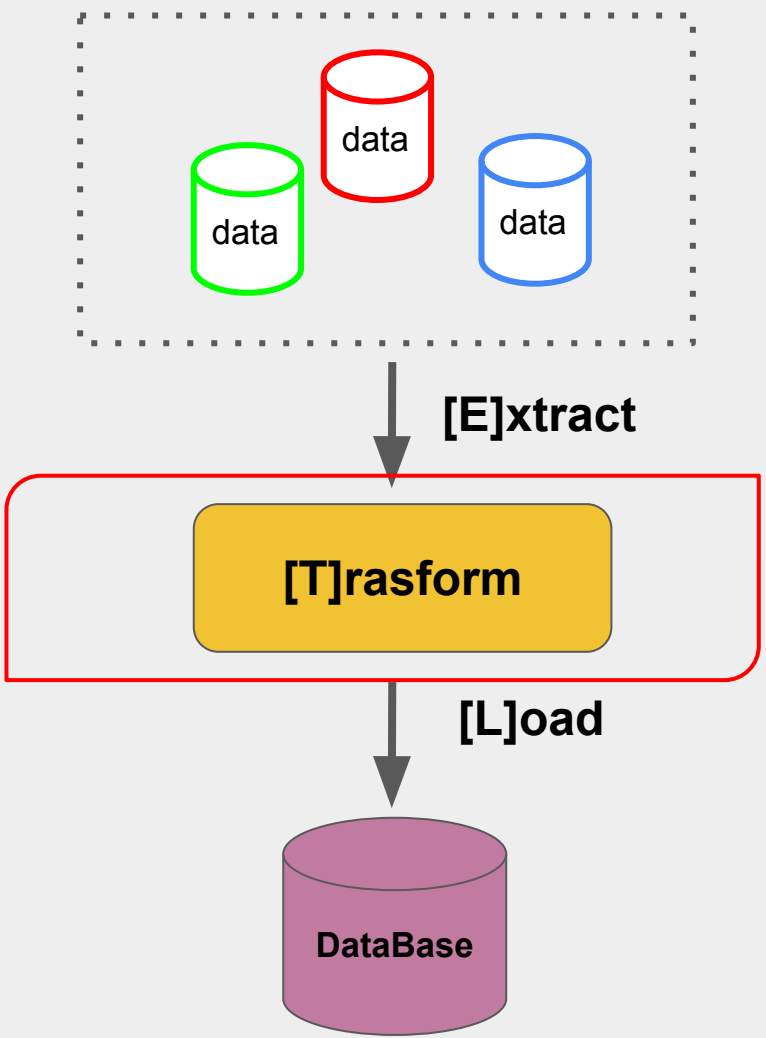
- **Database**
- **Fonti online e social media**
- **Dati aziendali (ERP)**
- **Applicazioni di vendita e marketing**
- **Dati da flussi di transazioni**
- **Sistemi di gestione delle relazioni con i clienti (CRM)**
- **Dati rilasciati da dei sensori**



Open for Innovation
KNIME

ETL → **[T]**ransform(ation) = Trasformare

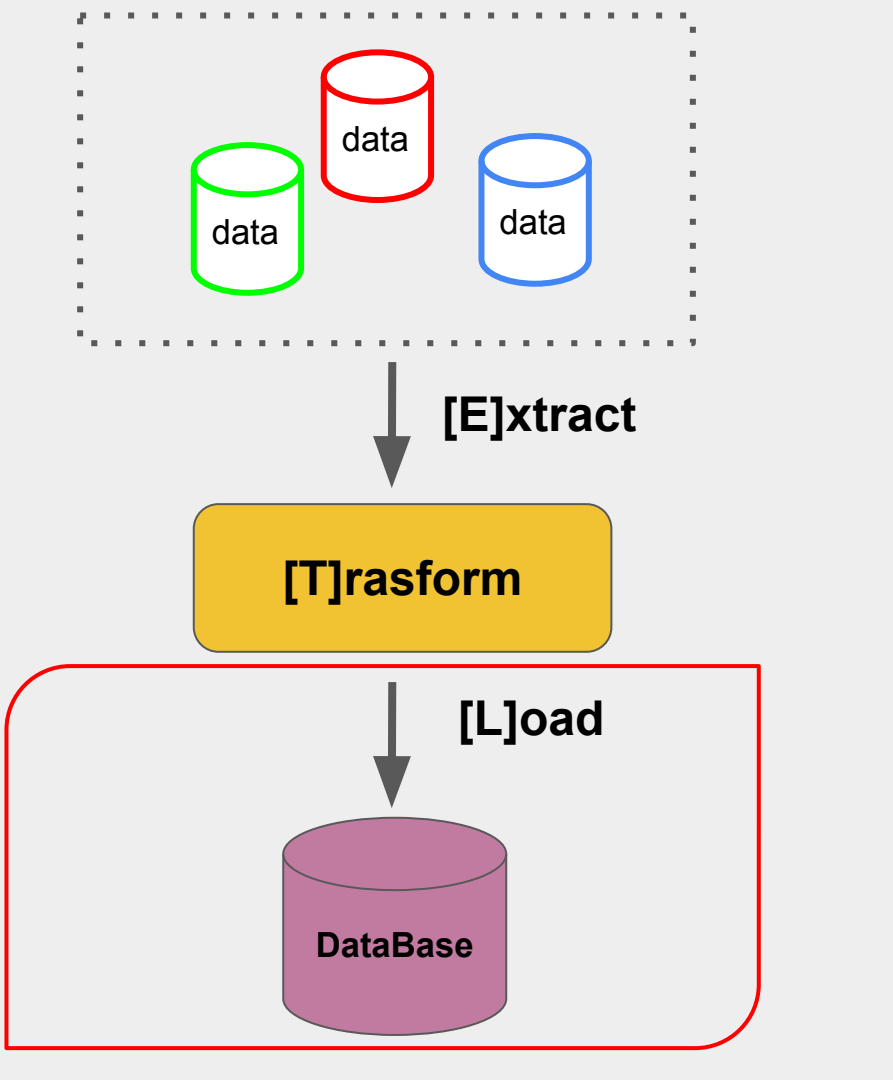
- **Standardizzazione dati**
- **Pulizia dati**
- **Eliminazione di ridondanza nei dati**
- **Controllo e verifica**



Open for Innovation
KNIME

ETL → [L]oad(ing) = Caricare

- **Caricamento in batch (completo)**
- **Caricamento incrementale per attivita' di Data Analitics**
- **Caricamento incrementale di allenamento per Machine Learning**



Progetto KNIME

Open for Innovation
KNIME

Il progetto **KNIME** nasce nel 2004 all'Università di Costanza in Germania.

La **K** iniziale deriva da **Konstanz.**
KNIME = Konstanz Information Miner

Inizialmente focalizzato sulla ricerca farmaceutica

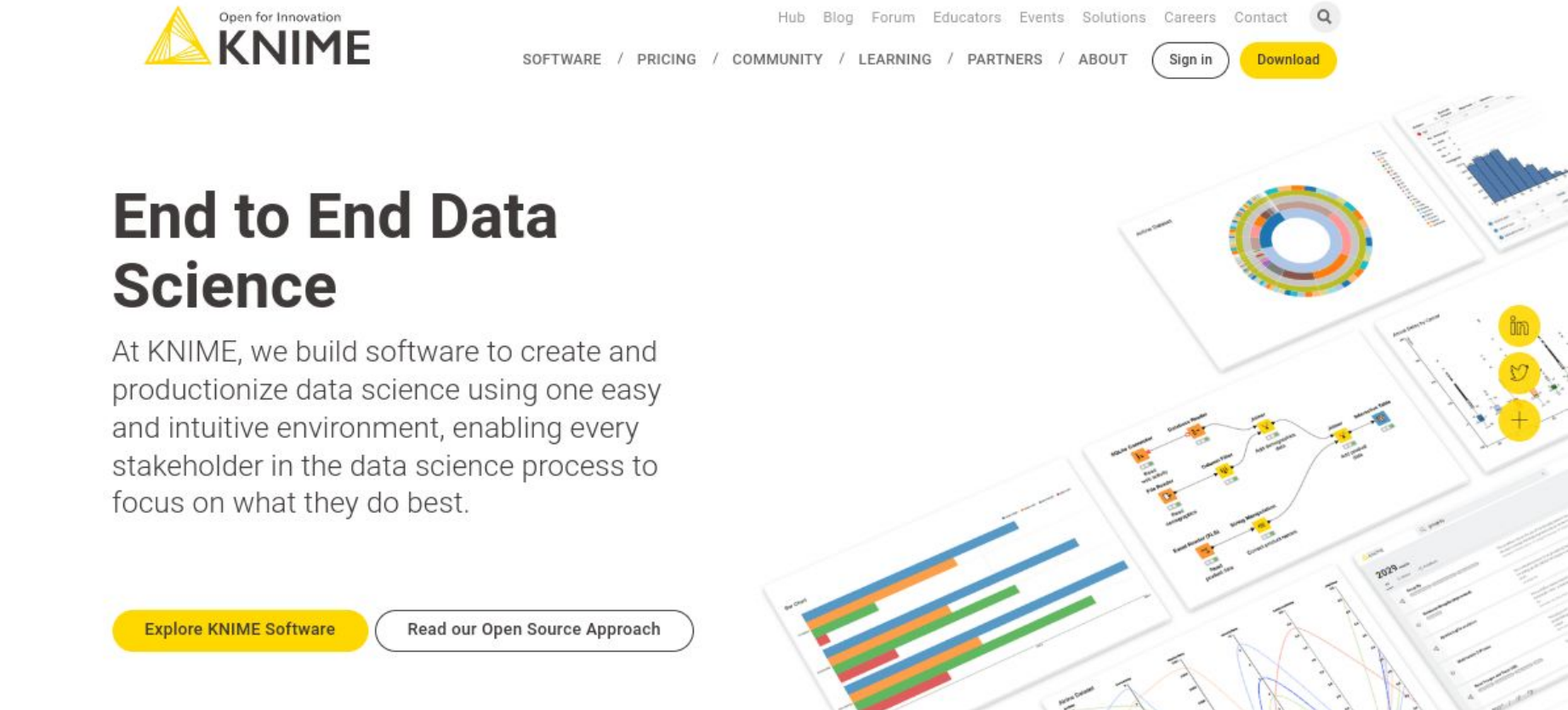
Universität Konstanz

Dove cercare KNIME ?

Dove cercare KNIME ?

in rete sul sito ufficiale www.knime.com

https://www.knime.com/

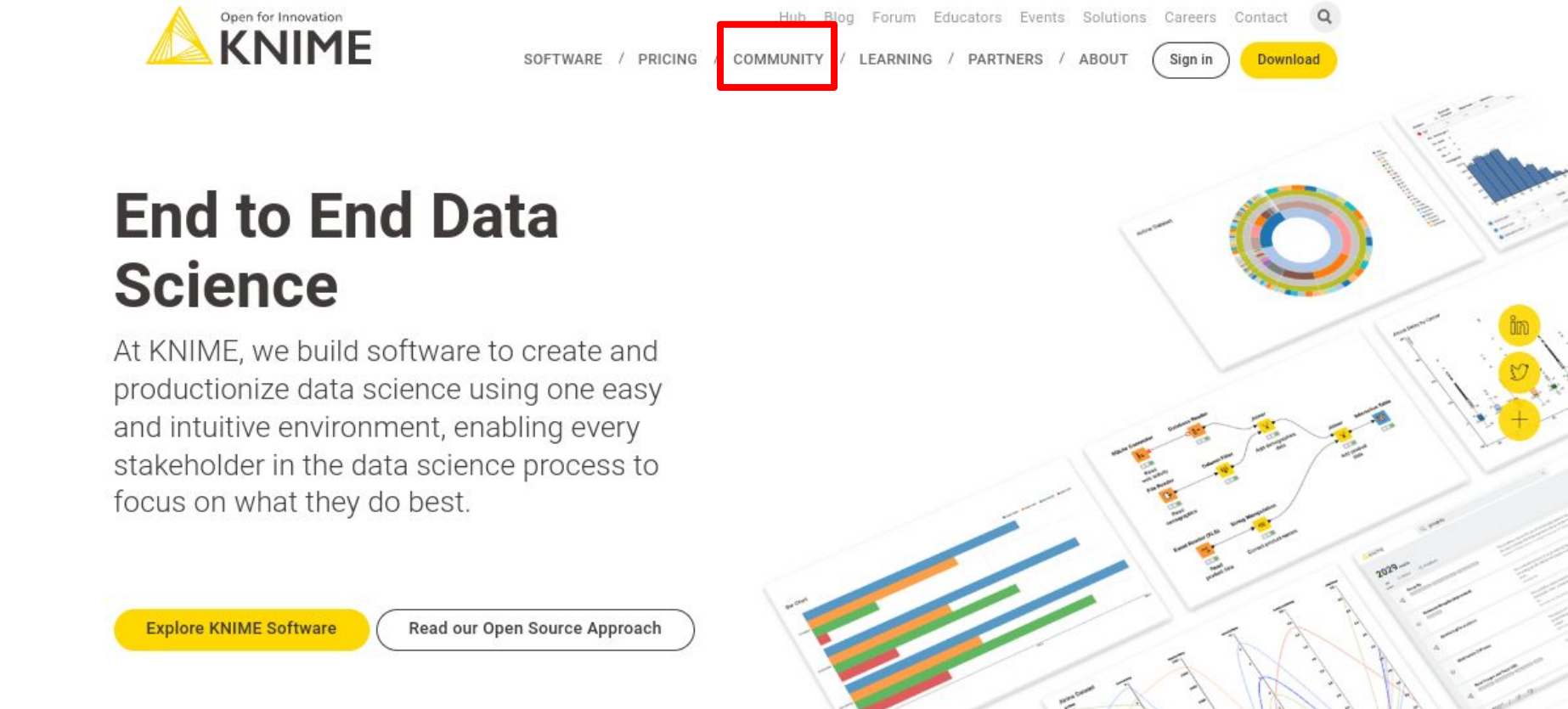


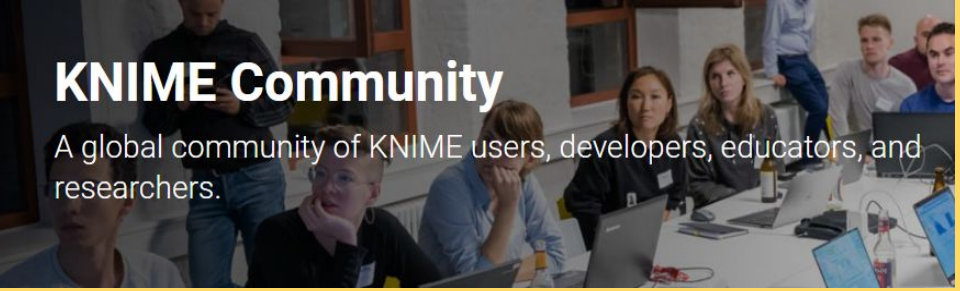
Contents:

- Community
- Learning
- Software
- Struttura, Nodi, Funzionamento

Community

https://www.knime.com/





KNIME Community

A global community of KNIME users, developers, educators, and researchers.

- **Users**
- **Extension Developers**
- **Educators**
- **Researchers**

fonte: https://www.knime.com/knime-community

Open for Innovation
KNIME

Users

See what the active community of data users and data professionals are doing with KNIME.

Extension Developers

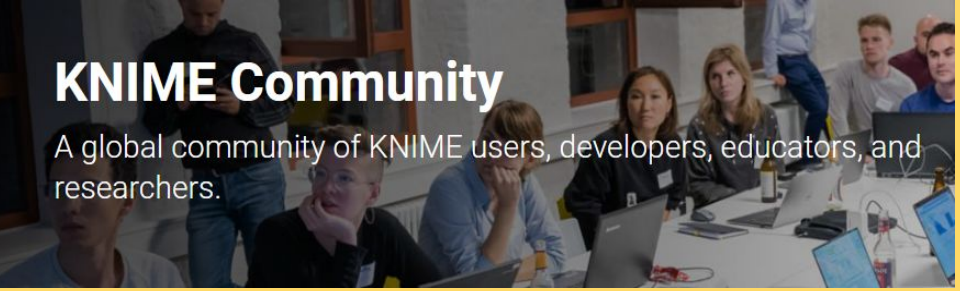
Create your own nodes and components and share with others on KNIME Hub.

Educators

Material and ideas for integrating KNIME in your data science classes.

Researchers

Read more about how scientists use KNIME in their research.



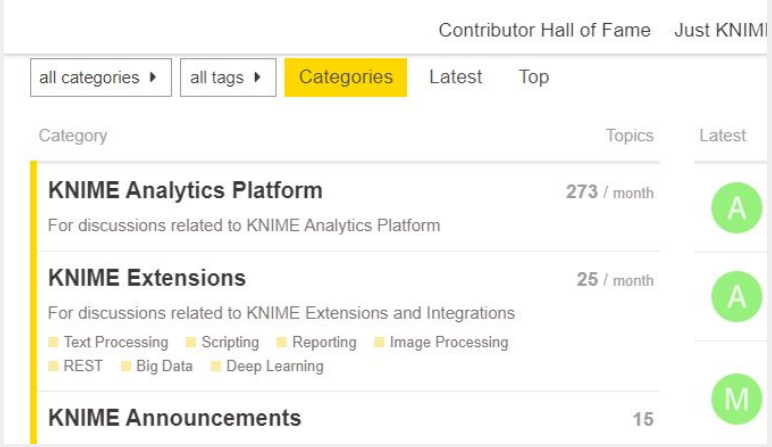
KNIME Community

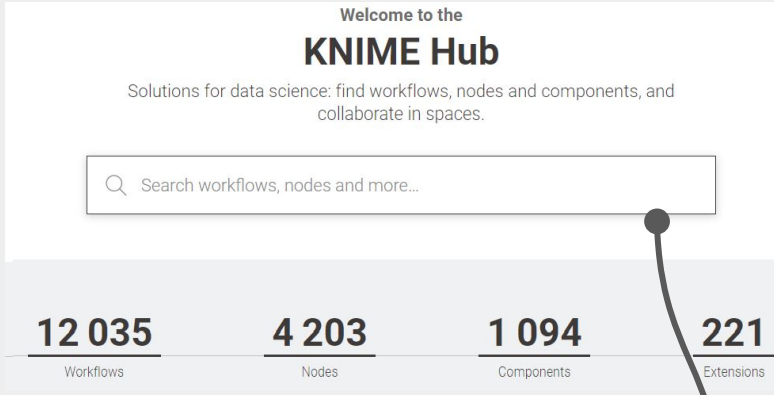
A global community of KNIME users, developers, educators, and researchers.

Become Part of the Global Community

Access knowledge and get help from data scientists, domain experts, other users, customers, and partners on both KNIME Forum and KNIME Hub. And, share your own knowledge and insights, too!

fonte: https://www.knime.com/knime-community





Open for Innovation
KNIME

Visit KNIME Forum

Visit KNIME Hub

- **KNIME Hub**

KNIME Hub

KNIME Hub is the place where you can search for nodes, extensions, components, and workflows made available by the community.

You can use these as blueprints and building blocks in KNIME Analytics Platform, and build workflows to solve your data science use cases.

KNIME Hub can also be used to upload and organize workflows in a central place, and to collaborate with other users on your projects.

fonte:https://docs.knime.com/latest/hub_user_guide/index.html#introduction

Via KNIME Hub you can:

- Find workflows, components uploaded by KNIME Community
- Upload and organize your workflows and components in projects
- Create and manage projects
- Collaborate with other users on your projects

https://hub.knime.com/

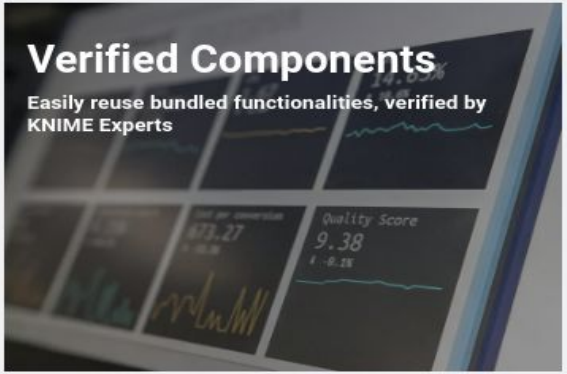
Welcome to the

KNIME Hub

Solutions for data science: find workflows, nodes and components, and collaborate in spaces.

Search workflows, nodes and more...

11 993	4 203	1 090	221
Workflows	Nodes	Components	Extensions



Ambiente simile a:

- **Rpubs** by RStudio

RPubs by RStudio

Sign in | Register

Getting Started with RPubs

RStudio lets you harness the power of R Markdown to create documents that weave together your writing and the output of your R code. And now, with RPubs, you can publish those documents on the web with the click of a button!

Prerequisites

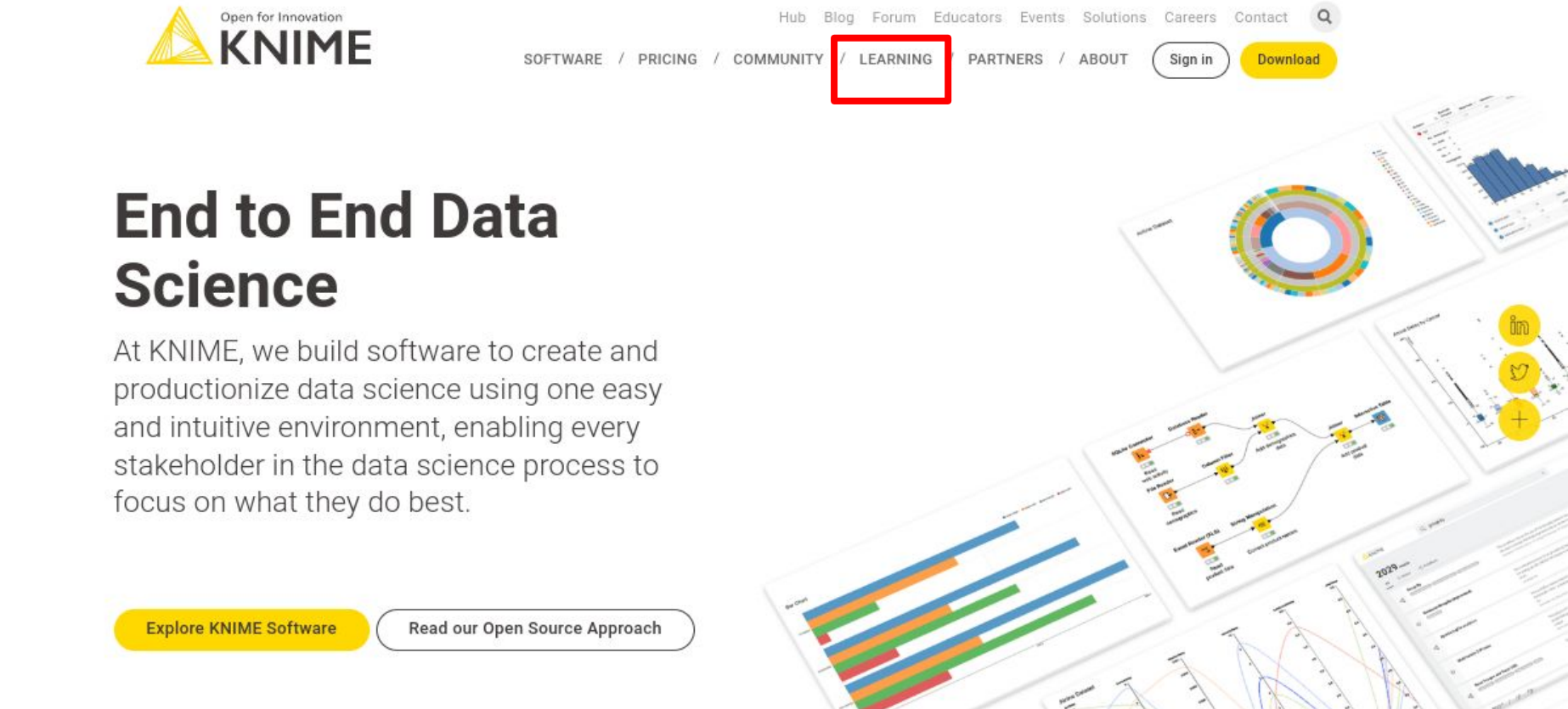
You'll need R itself, RStudio (v0.96.230 or later), and the knitr package (v0.5 or later).

Instructions

1. In RStudio, create a new R Markdown document by choosing `File | New | R Markdown`.
2. Click the `Knit HTML` button in the doc toolbar to preview your document.
3. In the preview window, click the `Publish` button.

Learning

https://www.knime.com/



Learning

Support material for getting started with KNIME: books, courses (online, onsite, and self-paced), technical documentation, certification, and more.

- **Getting Started Guide**
- **Software Documentation**
- **KNIME Hub**
- **KNIME Forum**

fonte:
https://www.knime.com/learning?utm_source=website&utm_medium=popup&utm_campaign=release

Open for Innovation
KNIME

Getting Started Guide

Find your way around the workbench, learn the traffic light system, start building your own workflow.

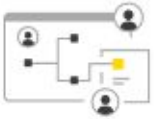
Get Started

Software Documentation

Read or download the technical documentation for KNIME Software.

Access Documentation

KNIME Hub



Find solutions for data science - workflows, nodes and components, and collaborate in spaces.

Visit Hub

KNIME Forum

Get answers to your data questions from the active, global community.

Visit Forum

Learning

Support material for getting started with KNIME: books, courses (online, onsite, and self-paced), technical documentation, certification, and more.

- **KNIME Courses**
- **Certified Trainers**
- **KNIME Certification**
- **Tutorials**

fonte:
https://www.knime.com/learning?utm_source=website&utm_medium=popup&utm_campaign=release

Open for Innovation
KNIME

KNIME Courses

Take a course - online, onsite, or self-paced - on a variety of topics.

Explore Courses

Certified Trainers

Trusted trainers endorsed by KNIME for their excellence as instructors.

Meet Trainers

KNIME Certification

Measure and certify your KNIME expertise.

Get Certified

Tutorials

Visit our YouTube channel for tutorials, webinar recordings, and user talks.

Visit Channel

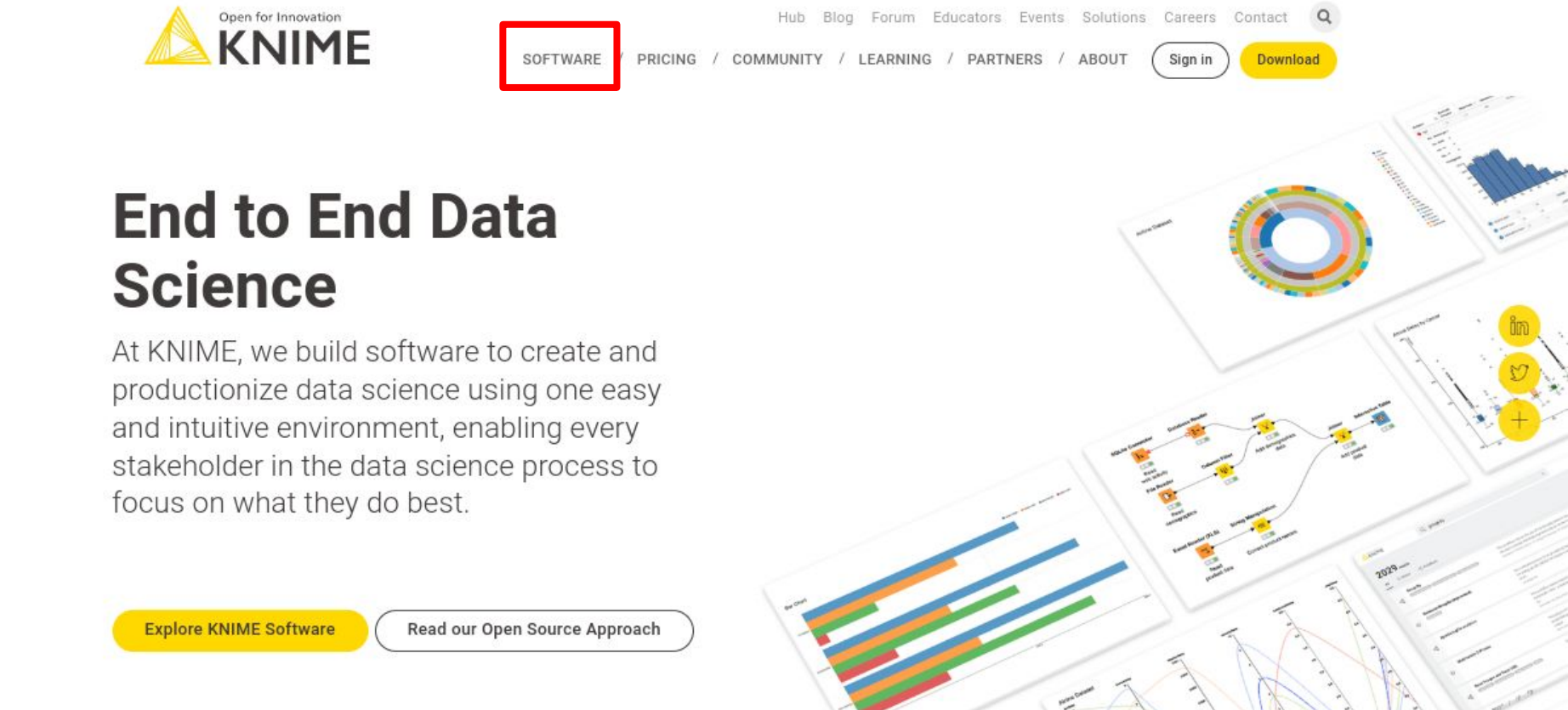
Open for Innovation
KNIME

filmato:

[1_communityLearning]

Software

https://www.knime.com/



KNIME Analytics Platform

KNIME Analytics Platform is the open source software for creating data science. Intuitive, open, and continuously integrating new developments, KNIME makes understanding data and designing data science workflows and reusable components accessible to everyone.

KNIME Server

KNIME Server is the enterprise software for team-based collaboration, automation, management, and deployment of data science workflows as analytical applications and services. Non experts are given access to data science via KNIME WebPortal or can use REST APIs.

fonte: https://www.knime.com/software-overview

KNIME Analytics Platform

KNIME Analytics Platform is the open source software for creating data science. Intuitive, open, and continuously integrating new developments, KNIME makes understanding data and designing data science workflows and reusable components accessible to everyone.

KNIME Server

KNIME Server is the enterprise software for team-based collaboration, automation, management, and deployment of data science workflows as analytical applications and services. Non experts are given access to data science via KNIME WebPortal or can use REST APIs.

fonte: https://www.knime.com/software-overview

KNIME Analytics Platform

Creating Data Science

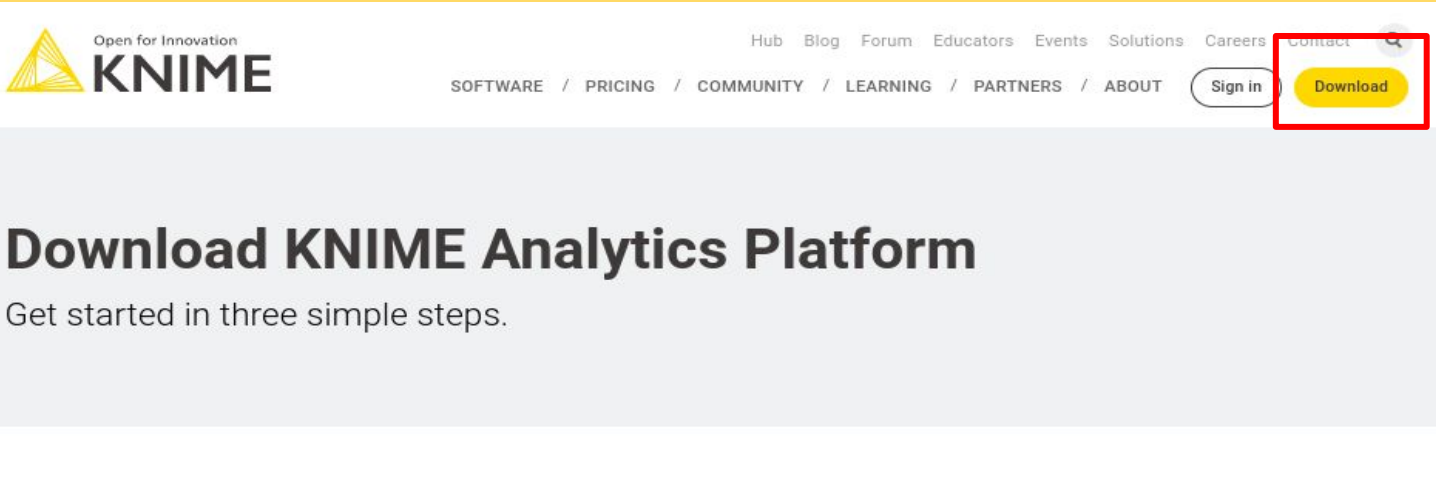
Download KNIME

KNIME Analytics Platform is the open source software for creating data science. Intuitive, open, and continuously integrating new developments, KNIME makes understanding data and designing data science workflows and reusable components accessible to everyone.

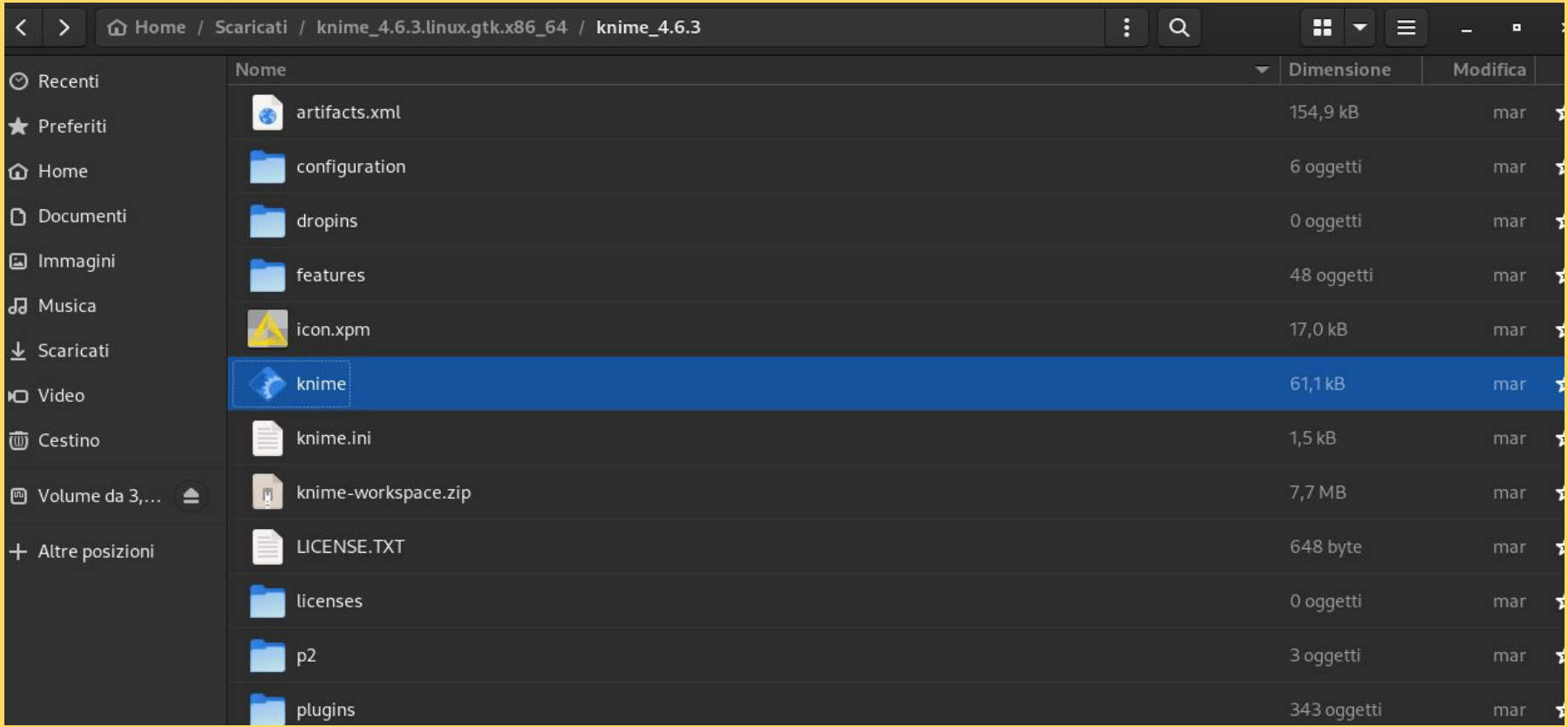
fonte: https://www.knime.com/knime-analytics-platform?utm_source=website&utm_medium=popup&utm_campaign=release

- Multipiattaforma:
  - Linux
  - Mac
  - Microsoft

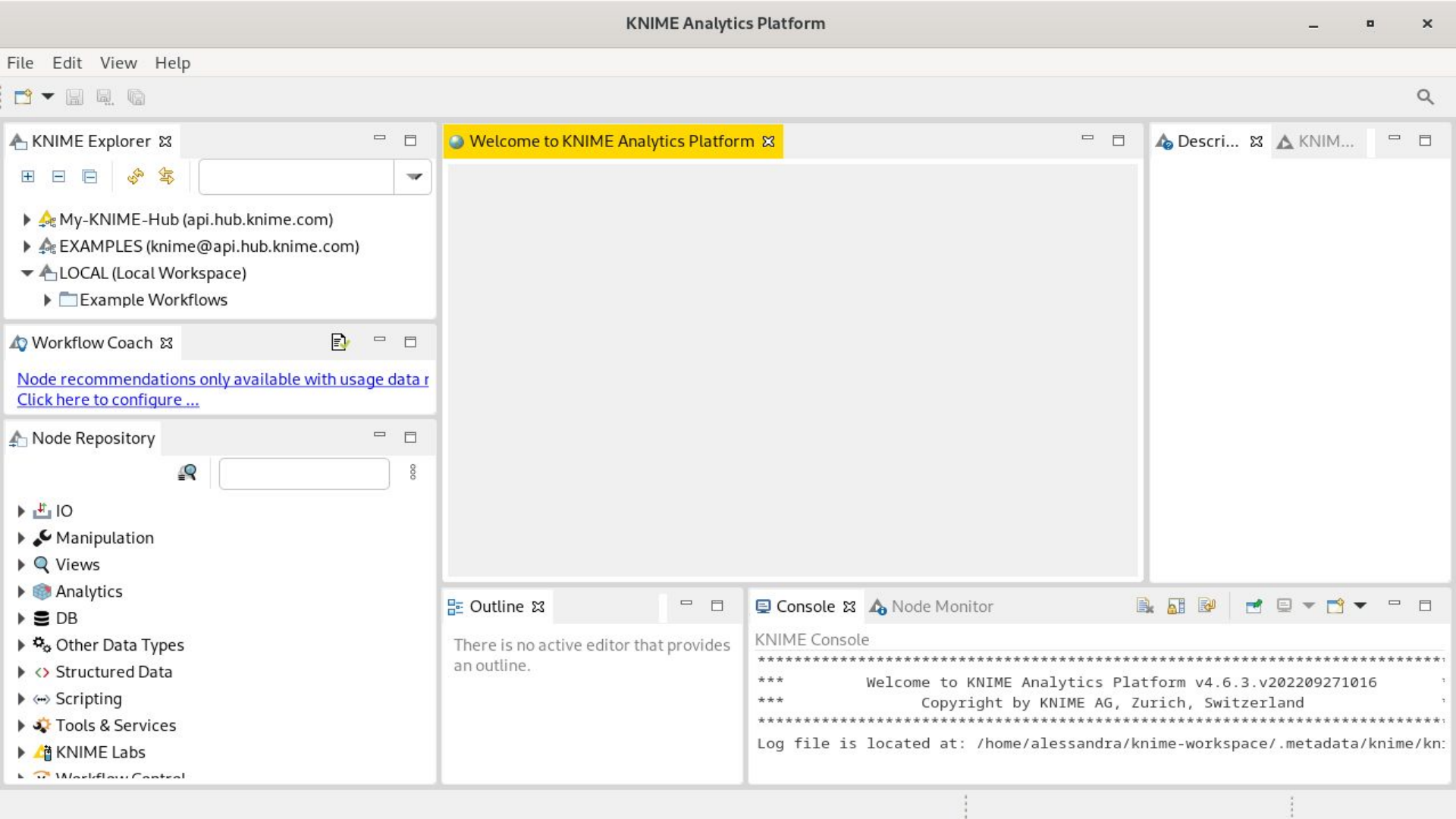
https://www.knime.com/ download



Installation(Linux)



Installation (Linux)


KNIME Analytics Platform
File
Edit
View
Help
KNIME Explorer
My-KNIME-Hub (api.hub.knime.com)
EXAMPLES (knime@api.hub.knime.com)
LOCAL (Local Workspace)
Example Workflows
Workflow Coach
Node recommendations only available with usage data r
Click here to configure ...
Node Repository
IO
Manipulation
Views
Analytics
DB
Other Data Types
Structured Data
Scripting
Tools & Services
KNIME Labs
Welcome to KNIME Analytics Platform
Descri...
KNIM...
Outline
There is no active editor that provides an outline.
Console
Node Monitor
KNIME Console
*** Welcome to KNIME Analytics Platform v4.6.3.v202209271016
*** Copyright by KNIME AG, Zurich, Switzerland
Log file is located at: /home/alessandra/knime-workspace/.metadata/knime/kn:

KNIME Analytics Platform

File Edit View Node Help

90%

KNIME Explorer

My-KNIME-Hub (api.hub.knime.com)
EXAMPLES (knime@api.hub.knime.com)
LOCAL (Local Workspace)
Example Workflows
2020_09_23_If_Else
KNIME_project1

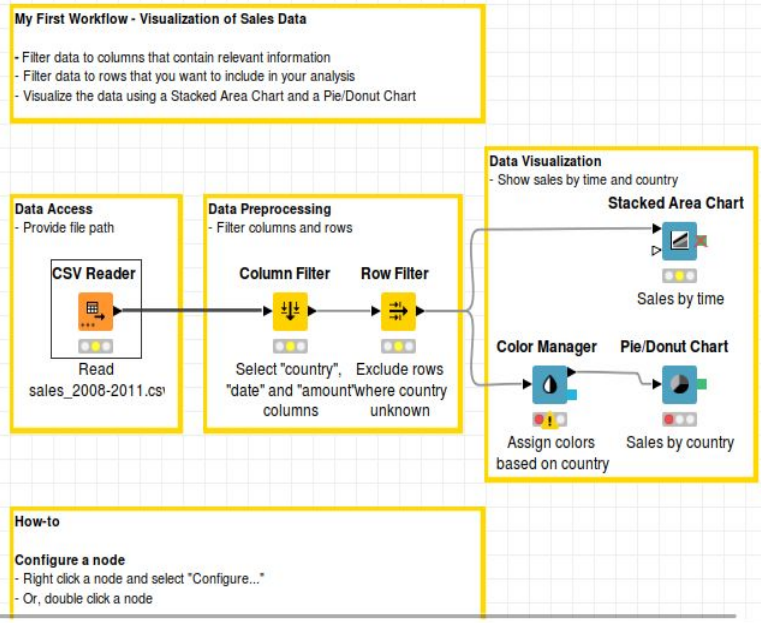
Workflow Coach

Node recommendations only available with usage data reporting.
Click here to configure ...

Node Repository

IO
Manipulation
Views
Analytics
DB
Other Data Types
Structured Data
Scripting
Tools & Services
KNIME Labs
Workflow Control
Workflow Abstraction

4: KNIME_project1 | 3: Visual Analysis of Sales Da... | Welcome to KNIME Analytics Pl...



Outline

Console | Node Monitor | Progress

KNIME Console
Log file is located at: /home/max/knime-workspace/.metadata/knime/knime.log

Description | KNIME Hub

CSV Reader

Reads CSV files. To auto-guess the structure of the file click the Autodetect format button. If you encounter problems with incorrect guessed data types disable the Limit data rows scanned option in the Advanced Settings tab. If the input file structure changes between different invocations, enable the Support changing file schemas option in the Advanced Settings tab. For further details see the KNIME File Handling Guide *File Handling Guide* .

Note: If you find that this node can't read your file, try the **File Reader** node. It offers more options for reading complex files.

This node can access a variety of different file systems. More information about file handling in KNIME can be found in the official File Handling Guide.

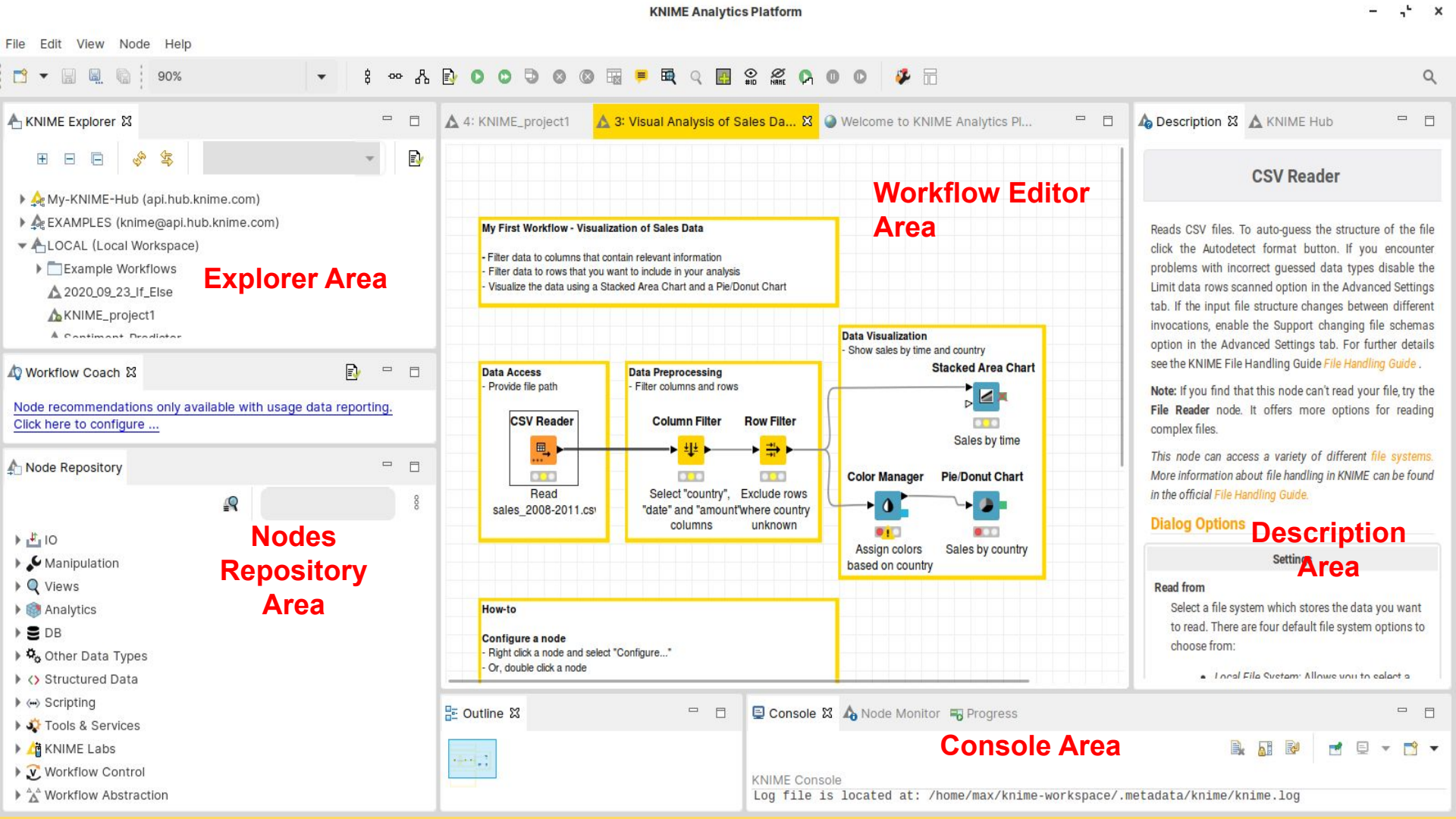
Dialog Options

Settings

Read from

Select a file system which stores the data you want to read. There are four default file system options to choose from:

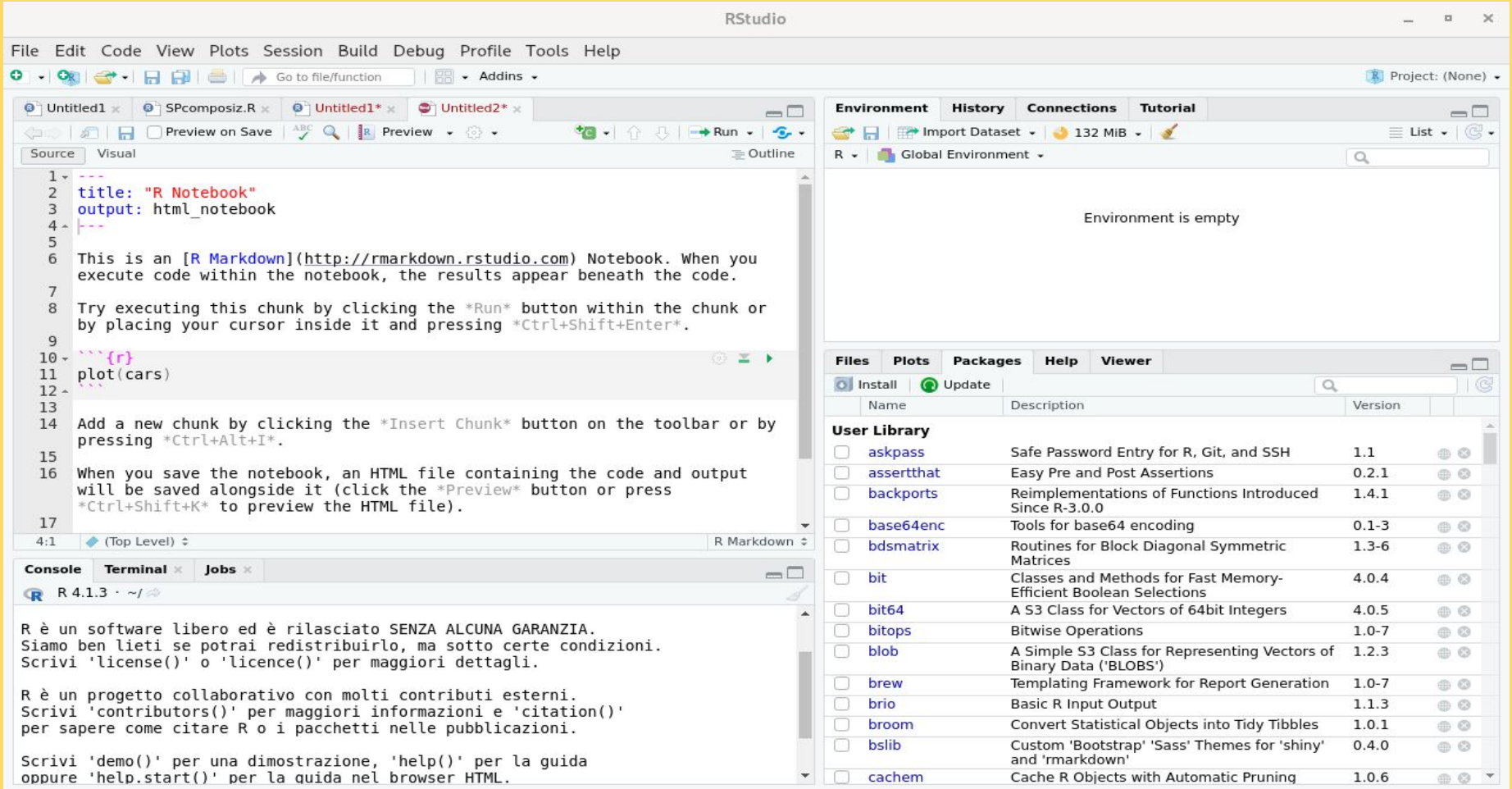
- *Local File System:* Allows you to select a


KNIME Analytics Platform
File Edit View Node Help
90%
KNIME Explorer
My-KNIME-Hub (api.hub.knime.com)
EXAMPLES (knime@api.hub.knime.com)
LOCAL (Local Workspace)
Example Workflows
2020_09_23_If_Else
KNIME_project1
Explorer Area
Workflow Coach
Node recommendations only available with usage data reporting.
Click here to configure ...
Node Repository
IO
Manipulation
Views
Analytics
DB
Other Data Types
Structured Data
Scripting
Tools & Services
KNIME Labs
Workflow Control
Workflow Abstraction
Nodes Repository Area
4: KNIME_project1
3: Visual Analysis of Sales Da...
Welcome to KNIME Analytics Pl...
Workflow Editor Area
My First Workflow - Visualization of Sales Data
- Filter data to columns that contain relevant information
- Filter data to rows that you want to include in your analysis
- Visualize the data using a Stacked Area Chart and a Pie/Donut Chart
Data Access
- Provide file path
CSV Reader
Read sales_2008-2011.csv
Data Preprocessing
- Filter columns and rows
Column Filter
Select "country", "date" and "amount" columns
Row Filter
Exclude rows where country unknown
Data Visualization
- Show sales by time and country
Stacked Area Chart
Sales by time
Color Manager
Assign colors based on country
Pie/Donut Chart
Sales by country
How-to
Configure a node
- Right click a node and select "Configure..."
- Or, double click a node
Description
KNIME Hub
CSV Reader
Reads CSV files. To auto-guess the structure of the file click the Autodetect format button. If you encounter problems with incorrect guessed data types disable the Limit data rows scanned option in the Advanced Settings tab. If the input file structure changes between different invocations, enable the Support changing file schemas option in the Advanced Settings tab. For further details see the KNIME File Handling Guide File Handling Guide .
Note: If you find that this node can't read your file, try the File Reader node. It offers more options for reading complex files.
This node can access a variety of different file systems. More information about file handling in KNIME can be found in the official File Handling Guide.
Dialog Options
Setting
Read from
Select a file system which stores the data you want to read. There are four default file system options to choose from:
Description Area
Outline
Console
Node Monitor
Progress
Console Area
KNIME Console
Log file is located at: /home/max/knime-workspace/.metadata/knime/knime.log

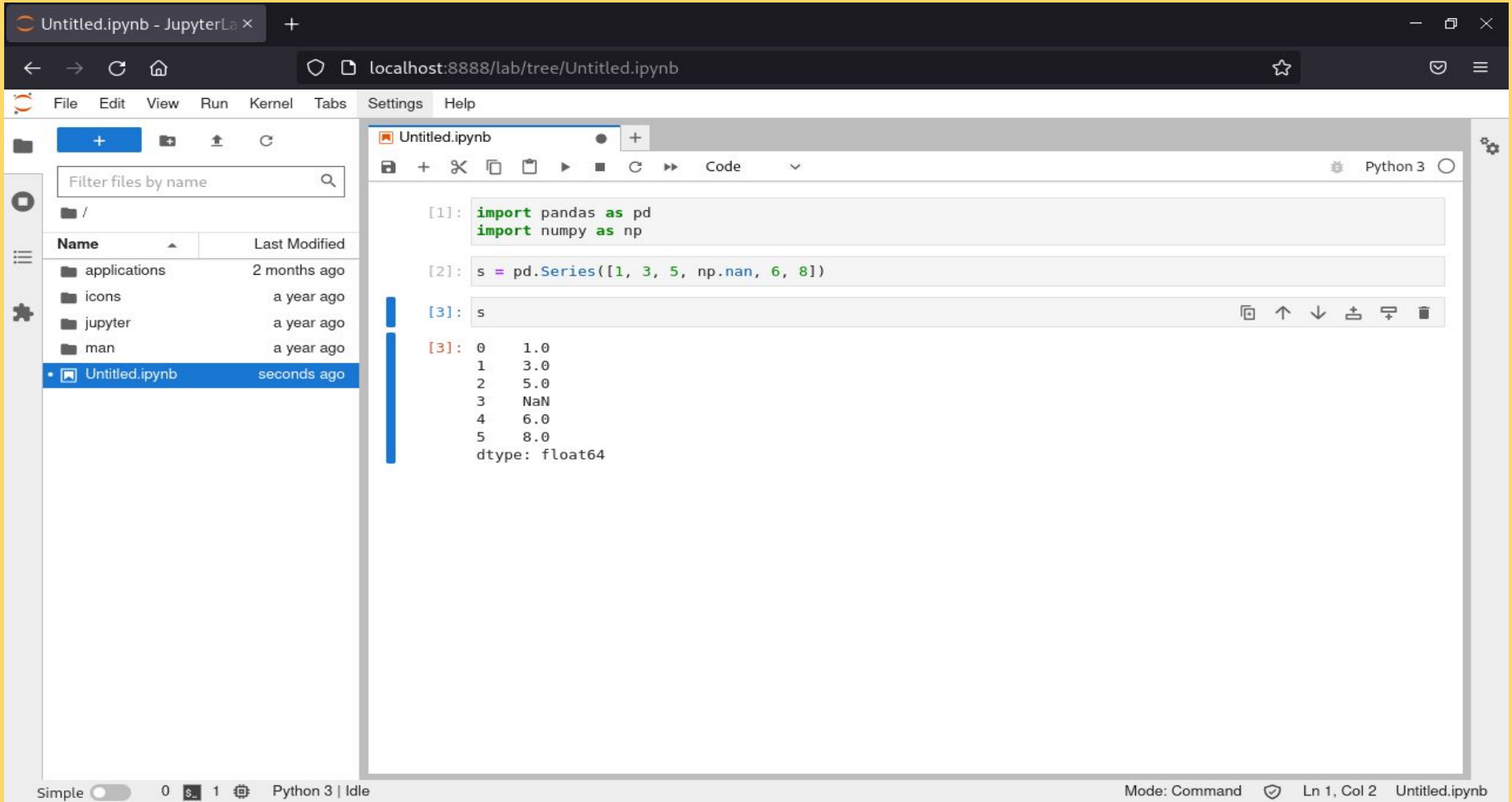
Confronto con altri ambienti di lavoro:

- RStudio
- Jupyter

RStudio


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Untitled1
SPcomposiz.R
Untitled1*
Untitled2*
Preview on Save
Preview
Run
Source
Visual
Outline
1 ---
2 title: "R Notebook"
3 output: html_notebook
4 ---
5
6 This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.
7
8 Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.
9
10 ```{r}
11 plot(cars)
12 ```
13
14 Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.
15
16 When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).
17
4:1
(Top Level)
R Markdown
Console
Terminal
Jobs
R 4.1.3 · ~/
R è un software libero ed è rilasciato SENZA ALCUNA GARANZIA.
Siamo ben lieti se potrai redistribuirlo, ma sotto certe condizioni.
Scrivi 'license()' o 'licence()' per maggiori dettagli.
R è un progetto collaborativo con molti contributi esterni.
Scrivi 'contributors()' per maggiori informazioni e 'citation()'
per sapere come citare R o i pacchetti nelle pubblicazioni.
Scrivi 'demo()' per una dimostrazione, 'help()' per la guida
oppure 'help.start()' per la guida nel browser HTML.
Environment
History
Connections
Tutorial
Import Dataset
132 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Install
Update
Name
Description
Version
User Library
askpass Safe Password Entry for R, Git, and SSH 1.1
assertthat Easy Pre and Post Assertions 0.2.1
backports Reimplementations of Functions Introduced Since R-3.0.0 1.4.1
base64enc Tools for base64 encoding 0.1-3
bdsmatrix Routines for Block Diagonal Symmetric Matrices 1.3-6
bit Classes and Methods for Fast Memory-Efficient Boolean Selections 4.0.4
bit64 A S3 Class for Vectors of 64bit Integers 4.0.5
bitops Bitwise Operations 1.0-7
blob A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS') 1.2.3
brew Templating Framework for Report Generation 1.0-7
brio Basic R Input Output 1.1.3
broom Convert Statistical Objects into Tidy Tibbles 1.0.1
bslib Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown' 0.4.0
cachem Cache R Objects with Automatic Pruning 1.0.6

Jupyter (Jupyter-lab)



Open for Innovation
KNIME

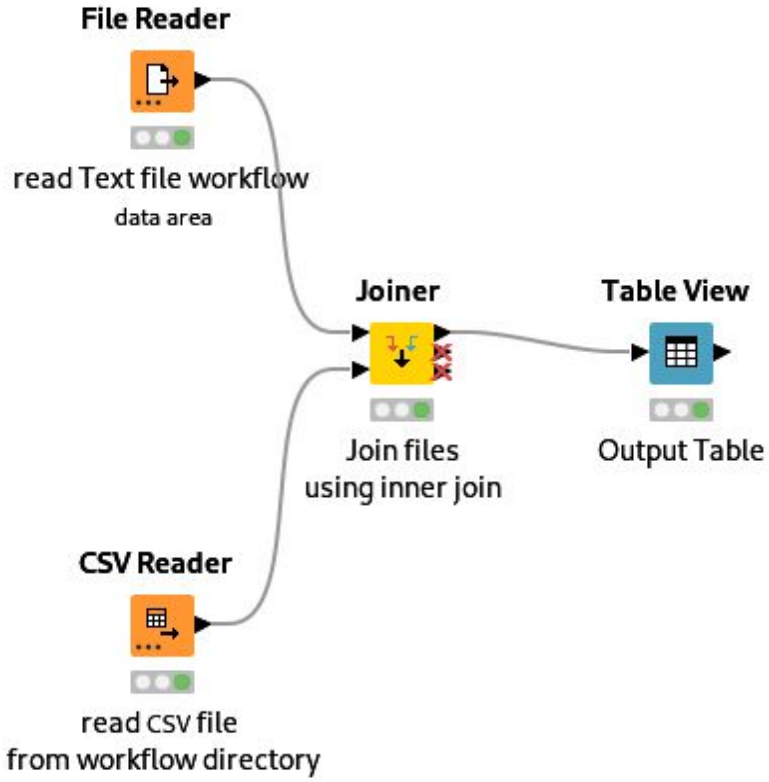
filmato:

2_download_avvioLinux

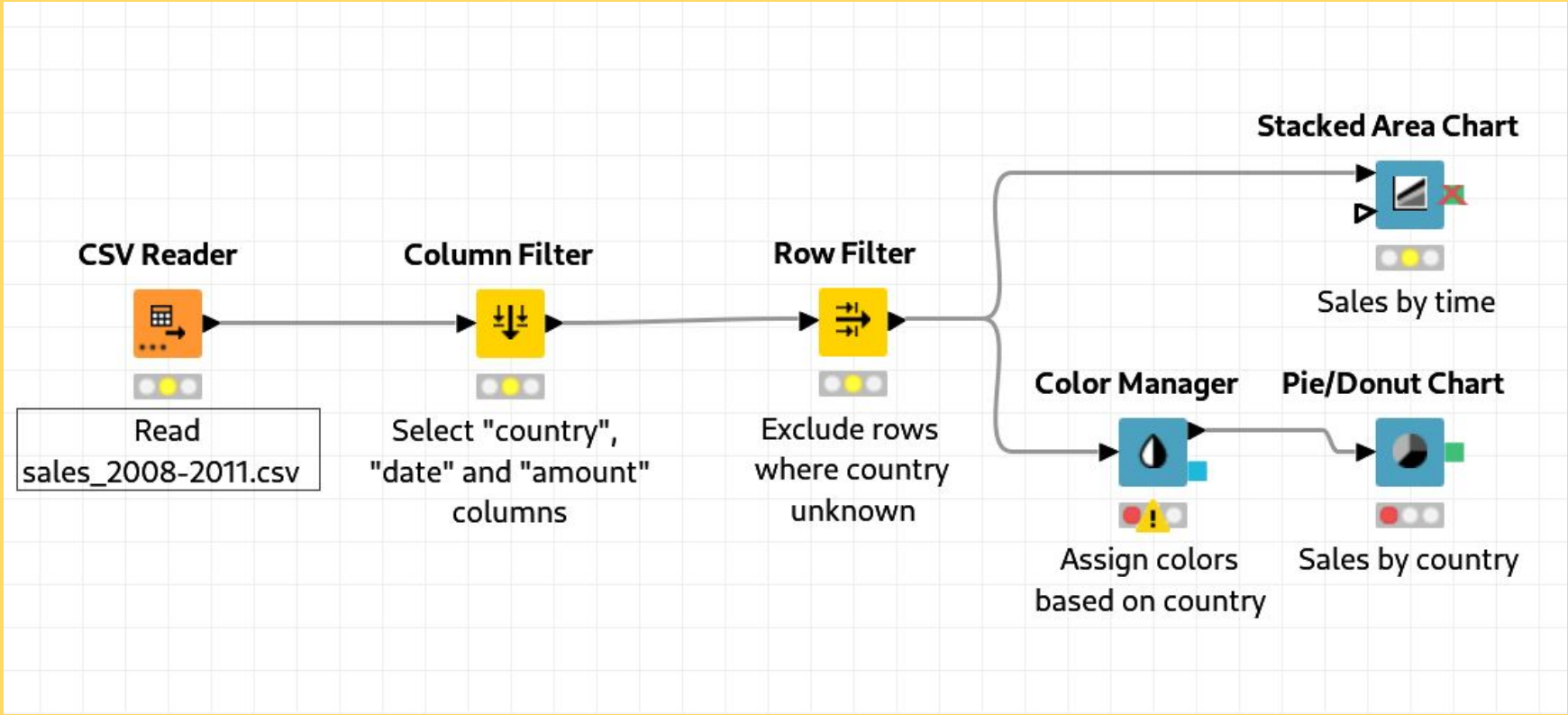
- **Struttura Logica (Workflow)**
- **Elementi (Nodes)**
- **Funzionamento**

Open for Innovation
KNIME

Strumento di analisi, reportistica, machine learning e data mining, basato su un **sistema di composizione di elementi grafici (Nodi) tra loro strutturalmente e logicamente interconnessi**.

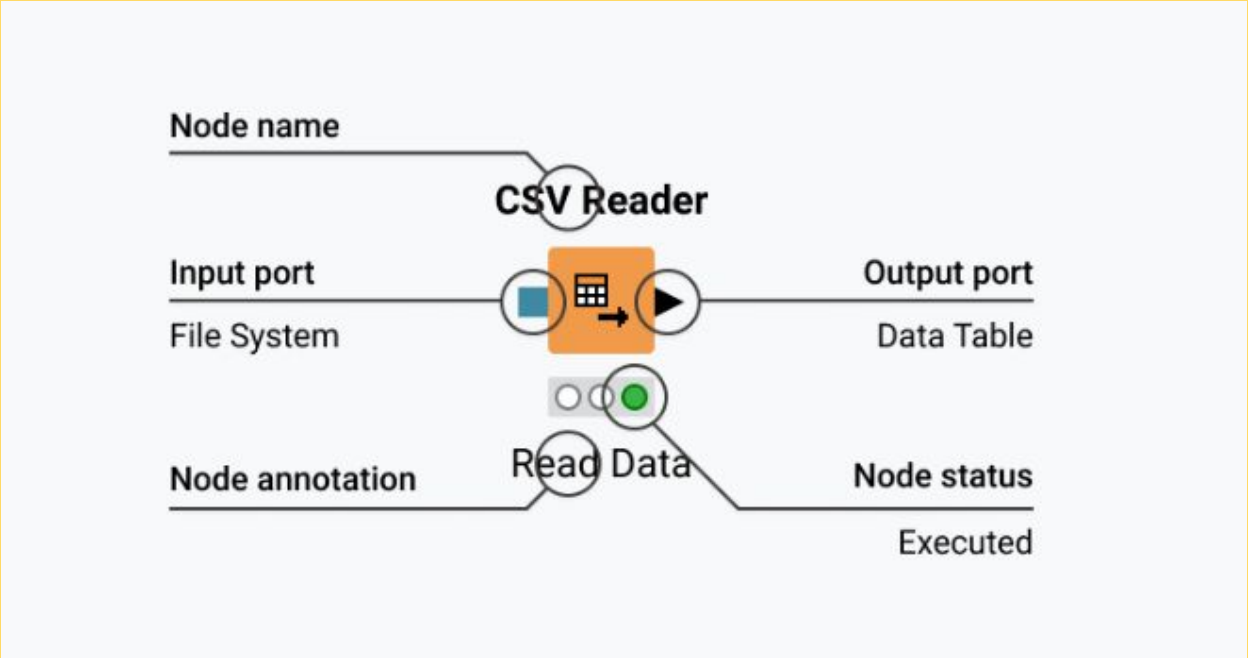


Struttura Logica (Workflow)



fonte: https://www.knime.com/getting-started-guide

Nodo



fonte: https://www.knime.com/getting-started-guide

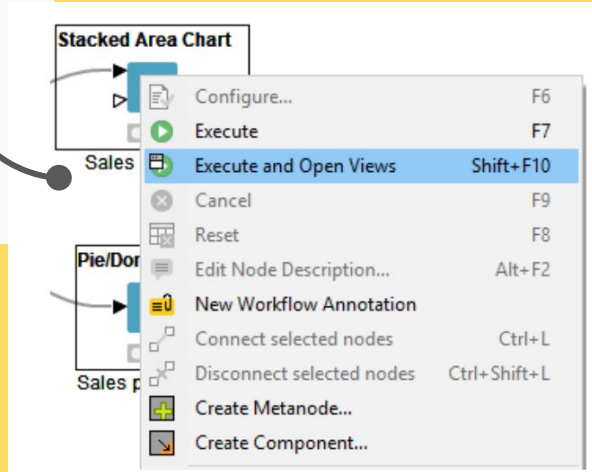
Nodo —> Semaforo degli stati

Not configured
The node is waiting for configuration or incoming data.

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

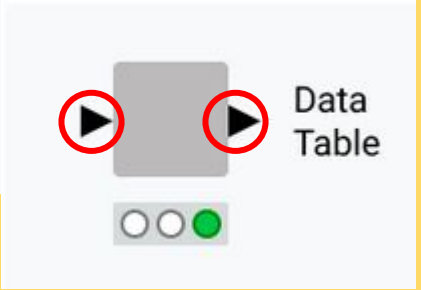
Executed
The node has been successfully executed. Results may be viewed and used in downstream nodes.

Error
The node has encountered an error during execution.



fonte: https://www.knime.com/getting-started-guide

Nodo —> Input output ports



fonte: https://www.knime.com/getting-started-guide

Open for Innovation
KNIME

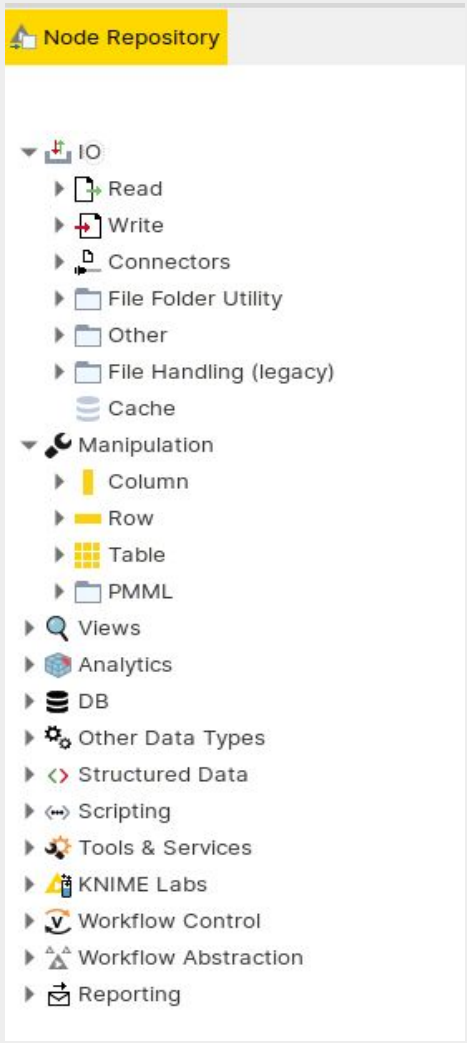
Node Repository Area

Organizzata a menu per categoria Nodi

La selezione dei Nodi si puo' fare con **doppio clic** o **drag and drop** sull'Area di lavoro.

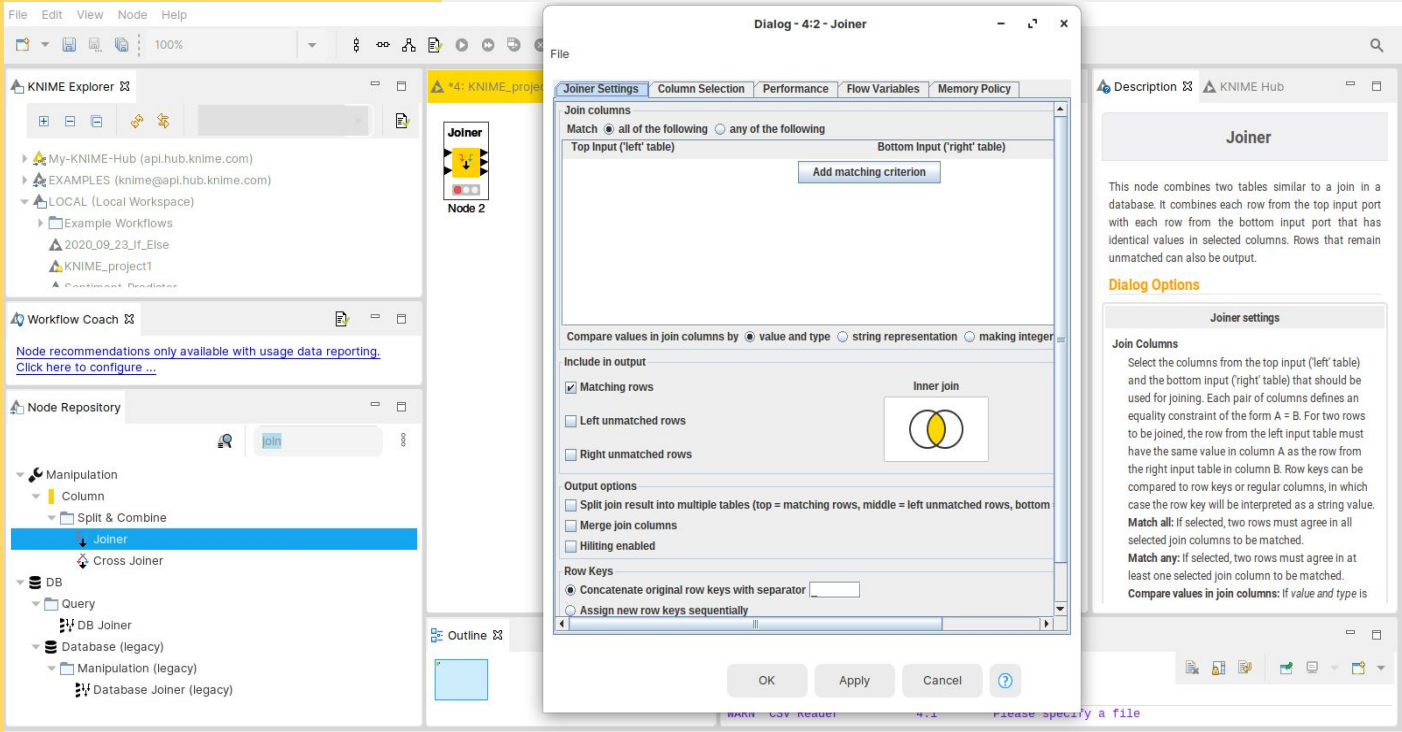
E' presente anche una **barra di ricerca**.

Per aggiungere Nodi con funzionalità diverse si accede a KNIME Hub.



Open for Innovation
KNIME

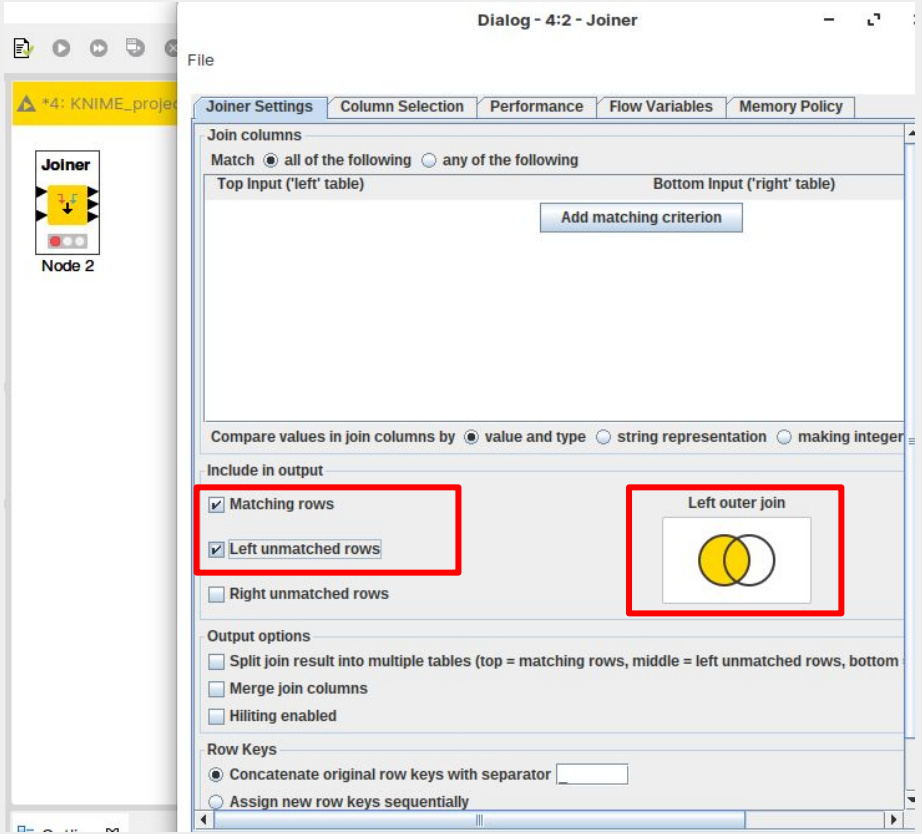
Entrando dentro ai nodi si aprono delle **maschere personalizzate di settaggio**, dove selezionare, escludere, trascinare
in modo visuale e grafico i parametri e gli attributi.



Open for Innovation
KNIME

Entrando dentro ai nodi si aprono delle **maschere personalizzate di settaggio**, dove selezionare, escludere, trascinare in modo visuale e grafico i parametri e gli attributi.

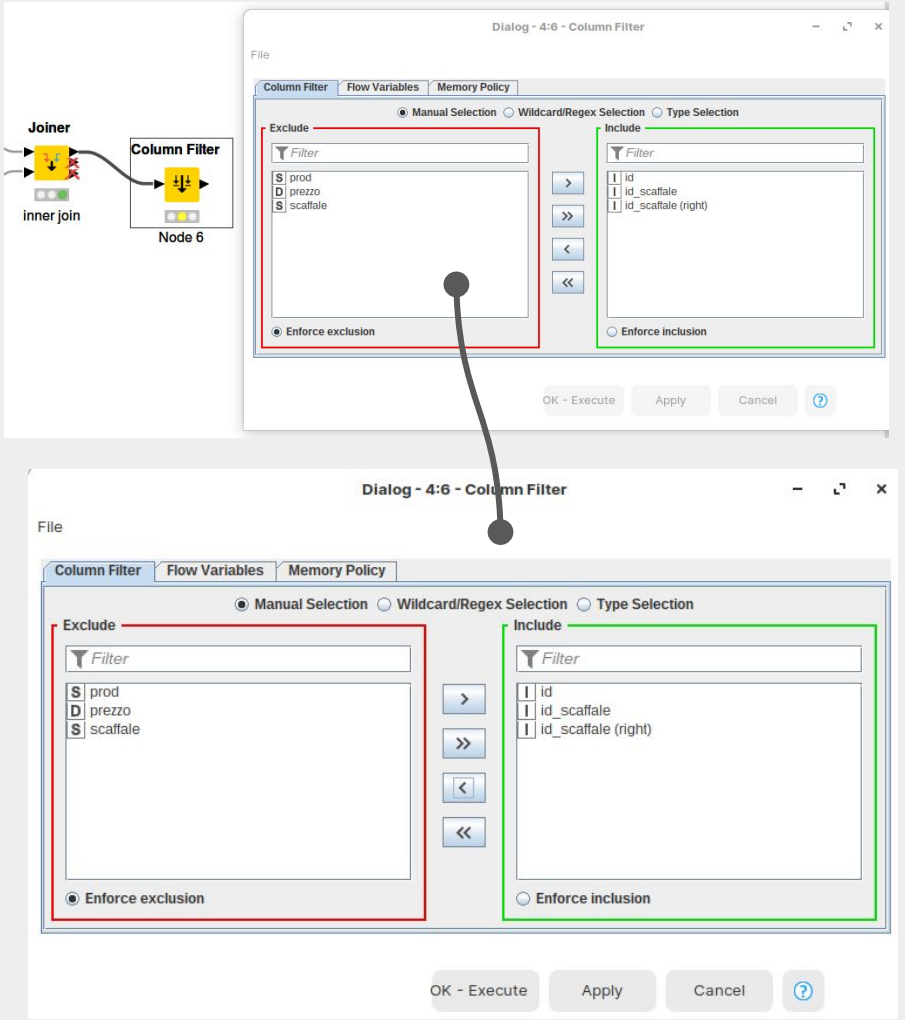
Esempio:
Nodo Join →Left outer Join



Open for Innovation
KNIME

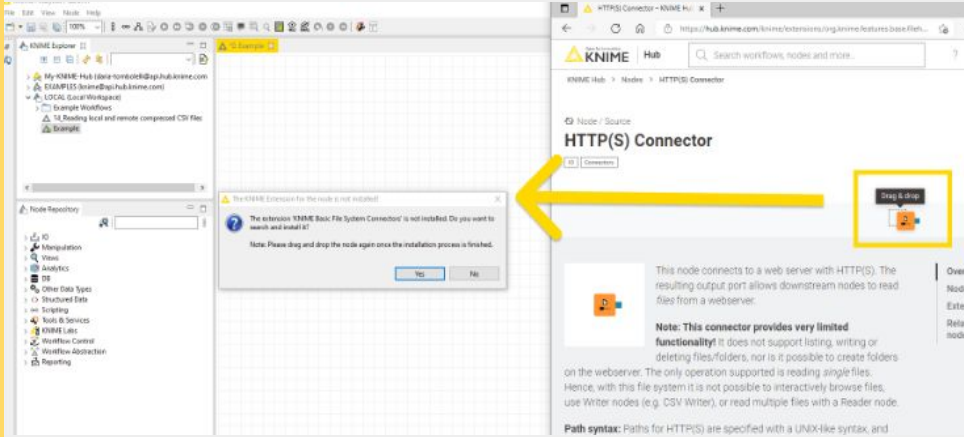
Entrando dentro ai nodi si aprono delle **maschere personalizzate di settaggio**, dove selezionare, escludere, trascinare in modo visuale e grafico i parametri e gli attributi.

Esempio:
Nodo Column Filter



KNIME Hub

You can **drag & drop** nodes, components, extensions, and workflows from KNIME Hub to import them into KNIME Analytics Platform and use them right away to build your own workflow, install KNIME Extensions, and execute uploaded workflow into your local installation.

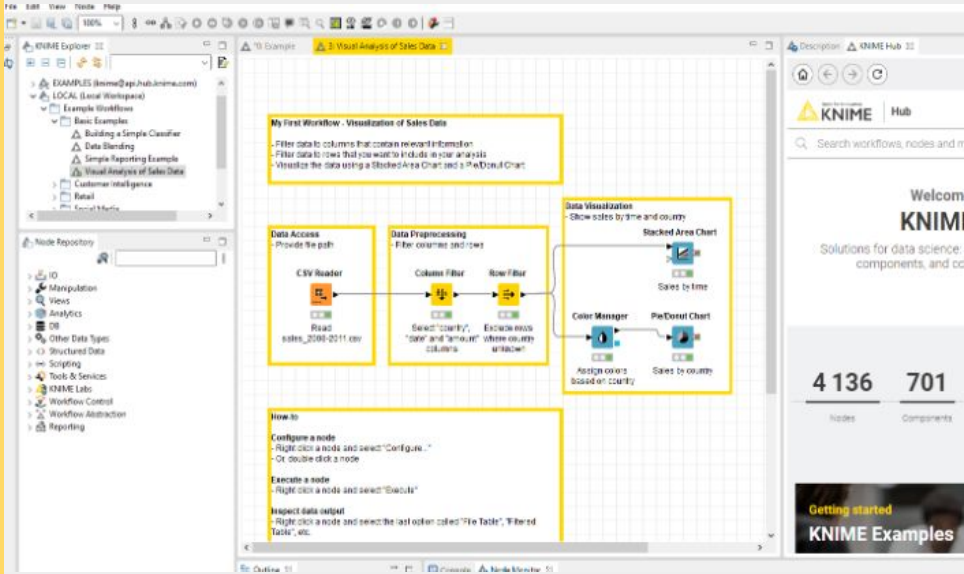


fonte:https://docs.knime.com/latest/hub_user_guide/index.html#introduction

Open for Innovation
KNIME

KNIME Hub

A KNIME Hub view is also available in **KNIME Analytics Platform**. It is simply an integrated browser in the KNIME Analytics Platform workbench so you can use KNIME Hub in the panel as you would do in an external browser window. Enter your search query, hit enter, and then drag & drop nodes, components, workflows, or extensions to the workbench of KNIME Analytics Platform.



fonte:https://docs.knime.com/latest/hub_user_guide/index.html#introduction

Open for Innovation
KNIME

https://hub.knime.com/

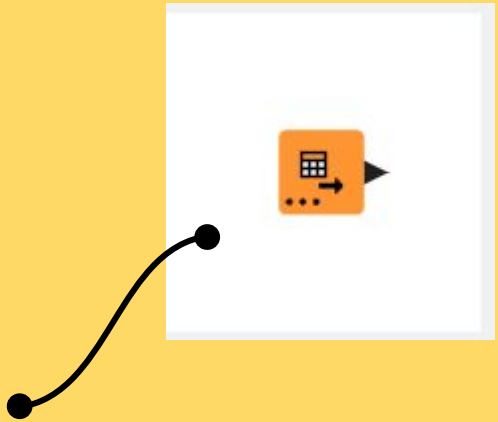
All | Workflows | Nodes | Components | Extensions

Node / Manipulator
Column Expressions
KNIME Labs | Streamable
This node provides the possibility to append an arbitrary number of columns or modify existing columns using expressions. For eac...

Node / Visualizer
Data Explorer
KNIME Labs | JavaScript Views (Labs)
The Data Explorer node offers a range of options for displaying properties of the input data in an interactive view. The node sup...

Node / Other
Write To Excel Template
Community Nodes | AF Utilities | ExcelUtil
This node writes the input data table into an existing Exc file without changing its formatting. Alternatively a new will ...

Node / Source
CSV Reader
IO | Read | Streamable
Reads CSV files. To auto-guess the structure of the file click the Autodetect format button. If you encounter problems with incor...

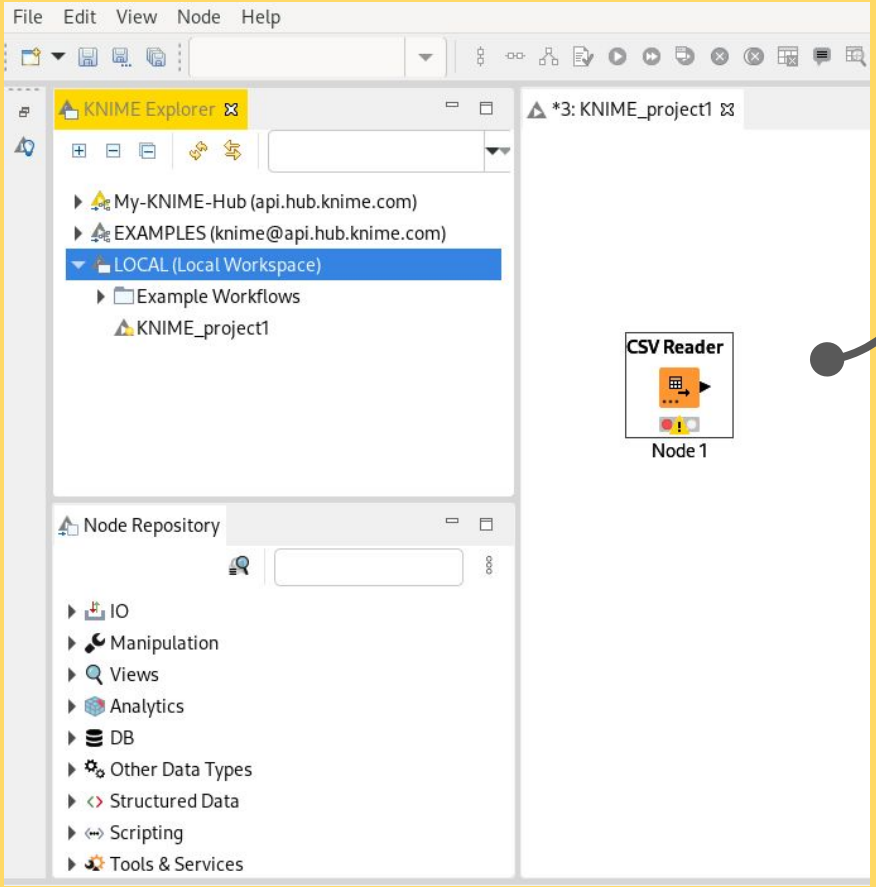


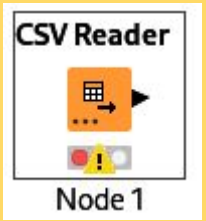
filmato:

[3.0_importWorkFlow]

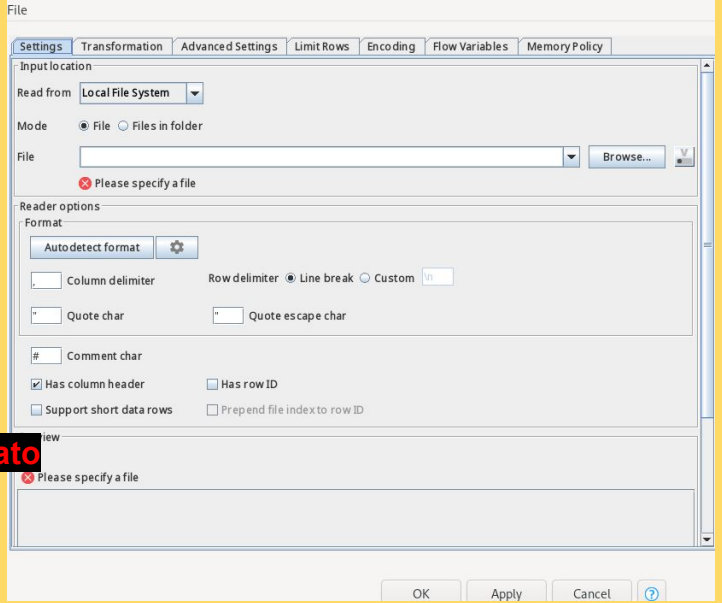
[3.1_importNodes]

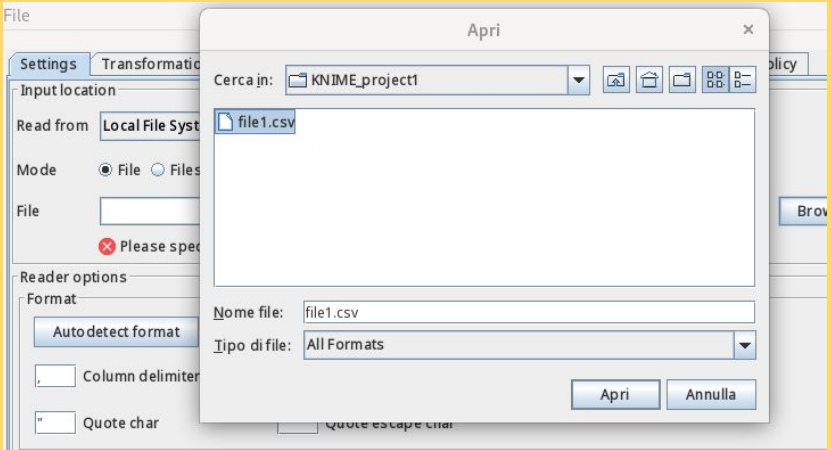
Setting Node CSV Reader



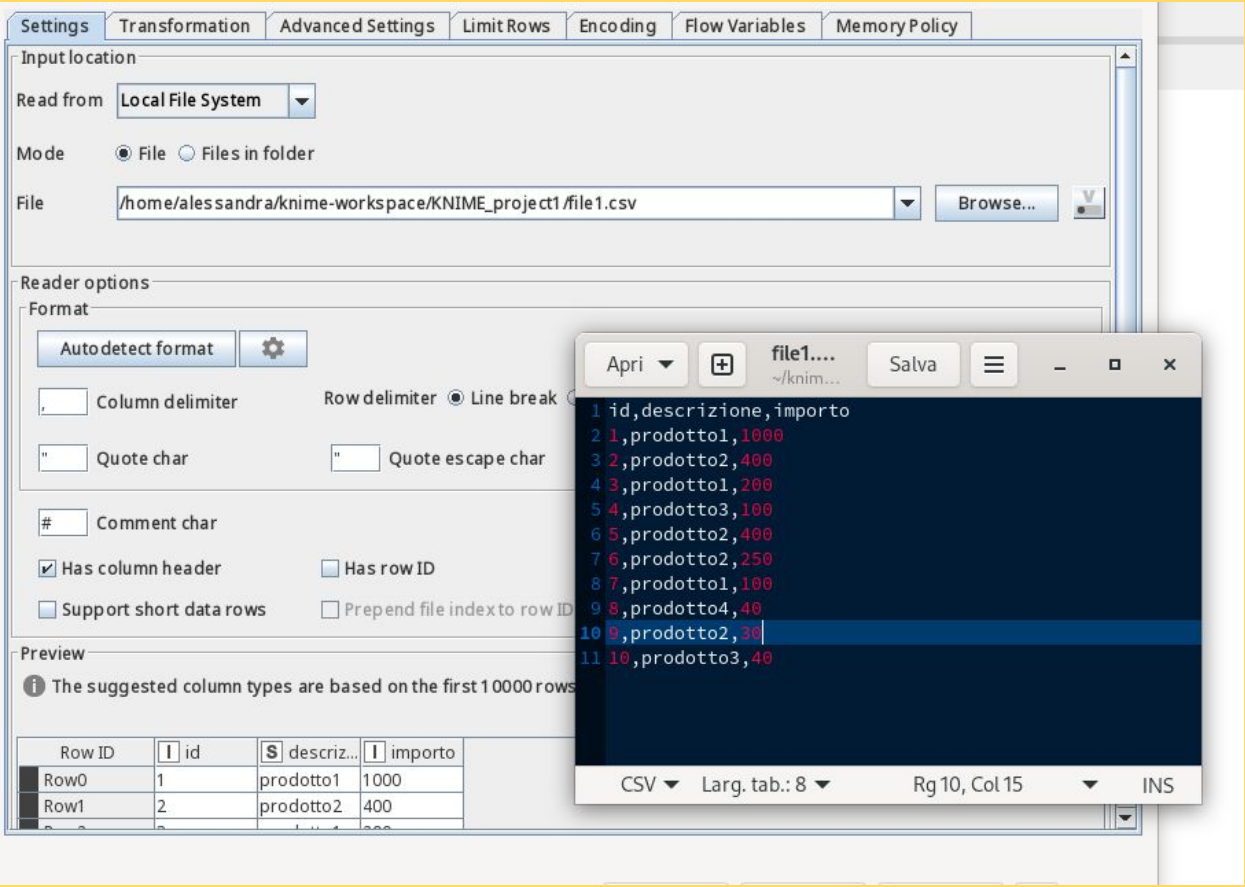


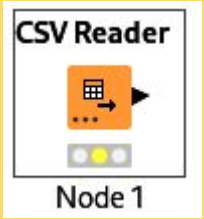
semaforo
non configurato
rosso





Execute Node CSV Reader

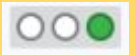




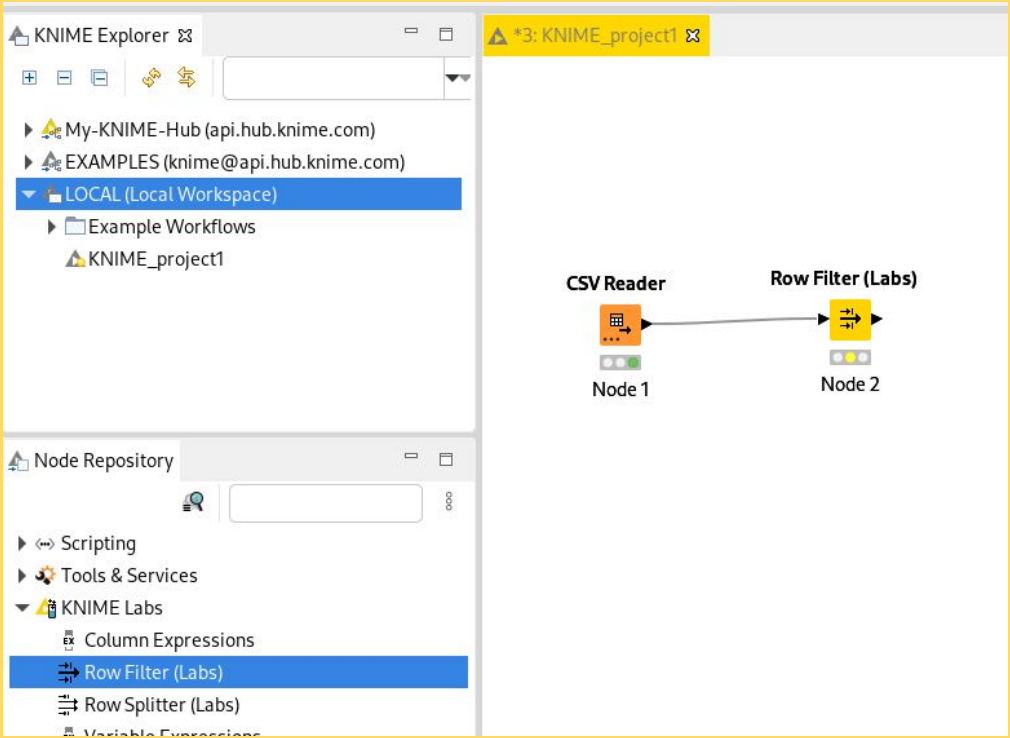
semaforo configurato giallo

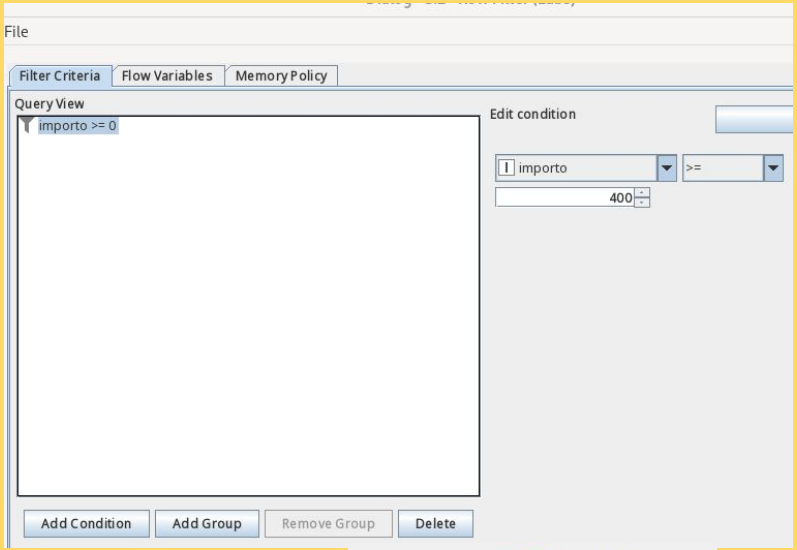
execute

semaforo eseguito verde

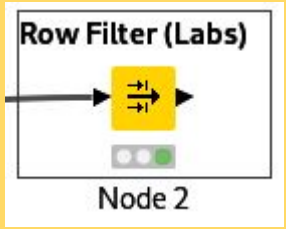


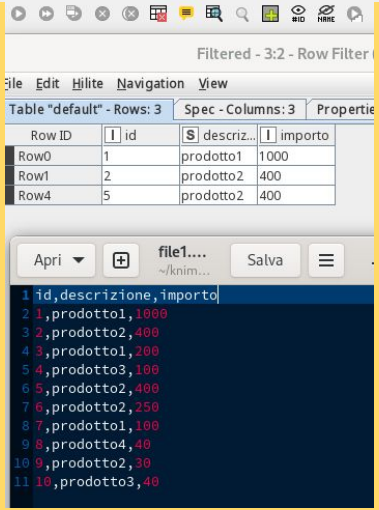
CSV Reader —> Row Filter(labs)





execute





esempi

Open for Innovation
KNIME

filmati:

4_exampleProject

5_examplejoin

6_example_filter

7_exampleSalesData

grazie per l’attenzione :)

Alessandra Santi

Linux Day a Pisa 2022

22 ottobre 2022