



UNIVERSITÀ
DEGLI STUDI
DI TERAMO



Big Data Analytics

Introduzione a KNIME

Prof.ssa Romina Eramo

Università degli Studi di Teramo

Dipartimento di Scienze della Comunicazione

reramo@unite.it

Cosa sono gli strumenti di analisi dei dati?

- » Gli strumenti di analisi dei dati sono soluzioni software, applicazioni e piattaforme che semplificano e accelerano il processo di analisi di grandi quantità di dati
 - Consentono la business intelligence, l'analisi, visualizzazione dati e reporting per le aziende
- » KNIME si posiziona tra migliori strumenti di analisi dei dati per facilità d'uso, interfaccia grafica, open source, espandibilità, modularità

Esempi di strumenti per l'analisi dei dati

<https://www.qrcode-tiger.com/it/tools-for-data-analysis>



KNIME: un player completo nel panorama analytics

» KNIME “si fa valere tra i grandi”, distinguendosi per versatilità, facilità d’uso e potenza integrate

Aspetto	Punti Chiave
 Data Input / ETL	Estrazione, trasformazione e caricamento dati semplici e visuali. Processi ETL rapidi rispetto a R, Python o Excel.
 Machine Learning	+200 algoritmi preconfigurati. Integrazione diretta con R e Python. Copertura ampia per ogni settore.
 Data Visualization	Funzionalità base. Per visualizzazioni avanzate: Spotfire, PowerBI, Tableau, Excel.
 User Friendliness	Interfaccia “What You See Is What You Get”: intuitiva, visuale e no-code oriented.

Collaborazione e Comunità: la forza di KNIME

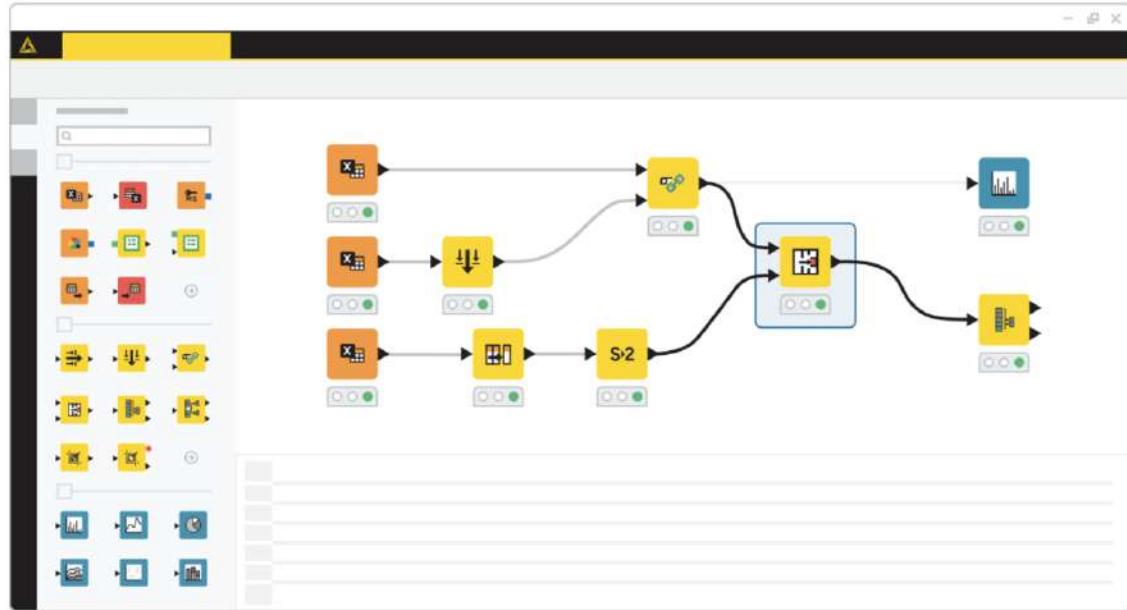
Aspetto	Punti Chiave
 Collaboration	Lavoro in team facilitato: workflow condivisibili, modulari, documentabili e riutilizzabili.
 Community Libraries	Ecosistema open source in crescita. Oltre 500.000 righe di codice sviluppate dalla community. Forum attivo e supporto costante.

» KNIME non è “il migliore in tutto”, ma eccelle per equilibrio: strumento versatile, autonomo e scalabile, capace di coprire la maggior parte delle esigenze analitiche.

Comparazione tra varie piattaforme per l'analisi dei dati



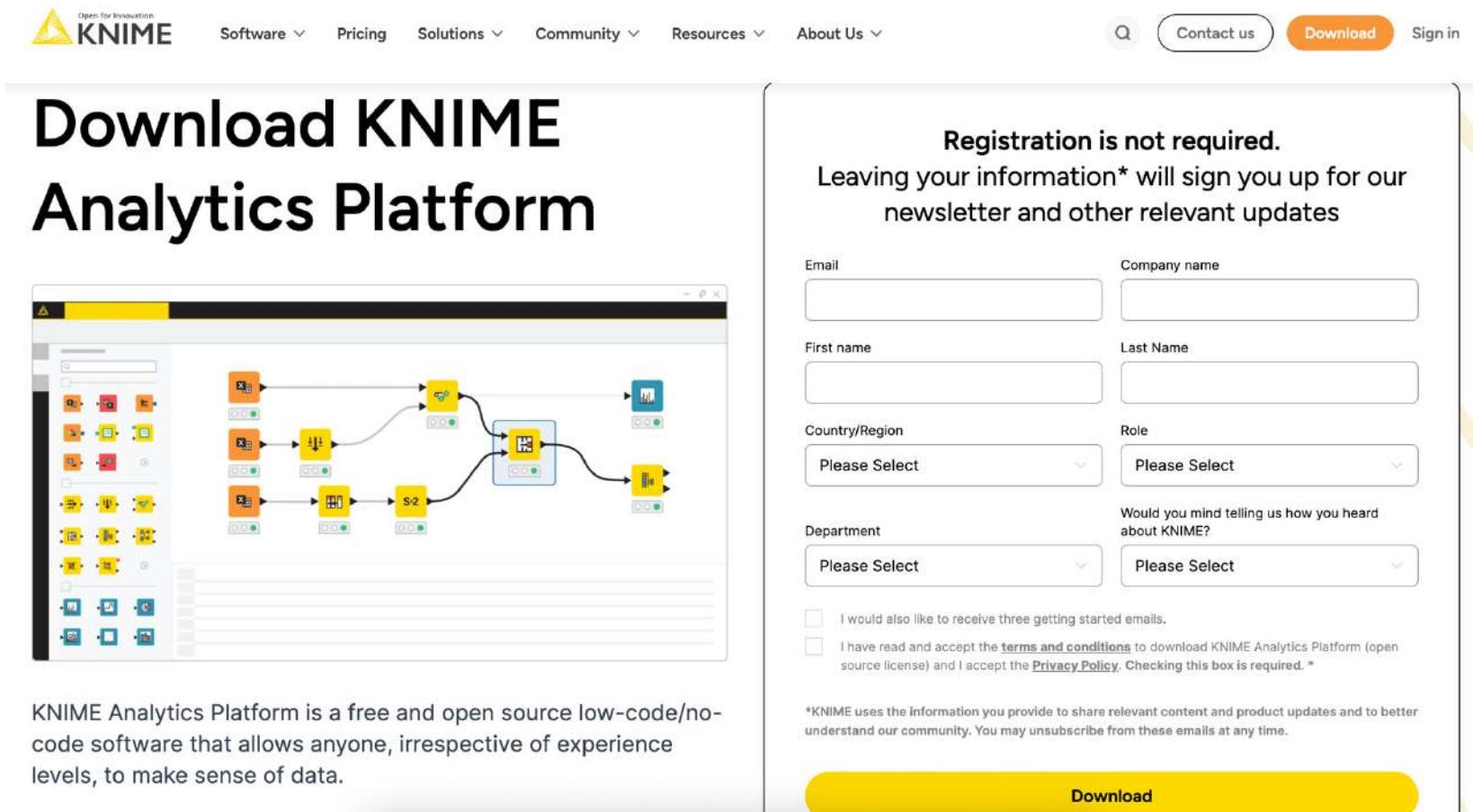
KNIME Analytics Platform



- » Software low-code/no-code gratuito e open source che permette a chiunque, indipendentemente dal livello di esperienza, di dare senso ai dati.
- » <https://www.knime.com/>

Scarica e installa

<https://www.knime.com/downloads>



The image shows the KNIME Analytics Platform download page. At the top, there is a navigation bar with links for Software, Pricing, Solutions, Community, Resources, About Us, a search bar, Contact us, Download (in an orange button), and Sign in. Below the navigation bar, there is a large heading "Download KNIME Analytics Platform" with a screenshot of the software interface showing a flow of nodes and data. A callout box on the right states "Registration is not required. Leaving your information* will sign you up for our newsletter and other relevant updates". It includes fields for Email, Company name, First name, Last name, Country/Region, Role, Department, and a question about how they heard about KNIME. There are also checkboxes for receiving emails and accepting terms and conditions. A yellow "Download" button is at the bottom.

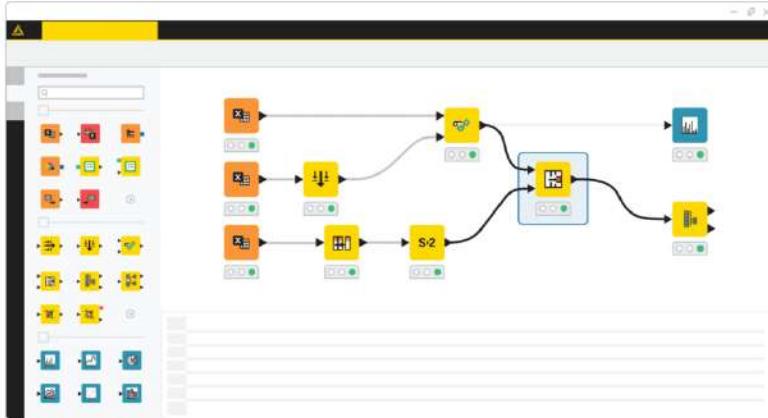
Open for Innovation

KNIME

Software ▾ Pricing Solutions ▾ Community ▾ Resources ▾ About Us ▾

Q Contact us Download Sign in

Download KNIME Analytics Platform



KNIME Analytics Platform is a free and open source low-code/no-code software that allows anyone, irrespective of experience levels, to make sense of data.

Registration is not required.
Leaving your information* will sign you up for our newsletter and other relevant updates

Email Company name

First name Last Name

Country/Region Role

Please Select Please Select

Department Would you mind telling us how you heard about KNIME?
Please Select Please Select

I would also like to receive three getting started emails.

I have read and accept the [terms and conditions](#) to download KNIME Analytics Platform (open source license) and I accept the [Privacy Policy](#). Checking this box is required. *

*KNIME uses the information you provide to share relevant content and product updates and to better understand our community. You may unsubscribe from these emails at any time.

Download

Muoversi in KNIME: i concetti fondamentali

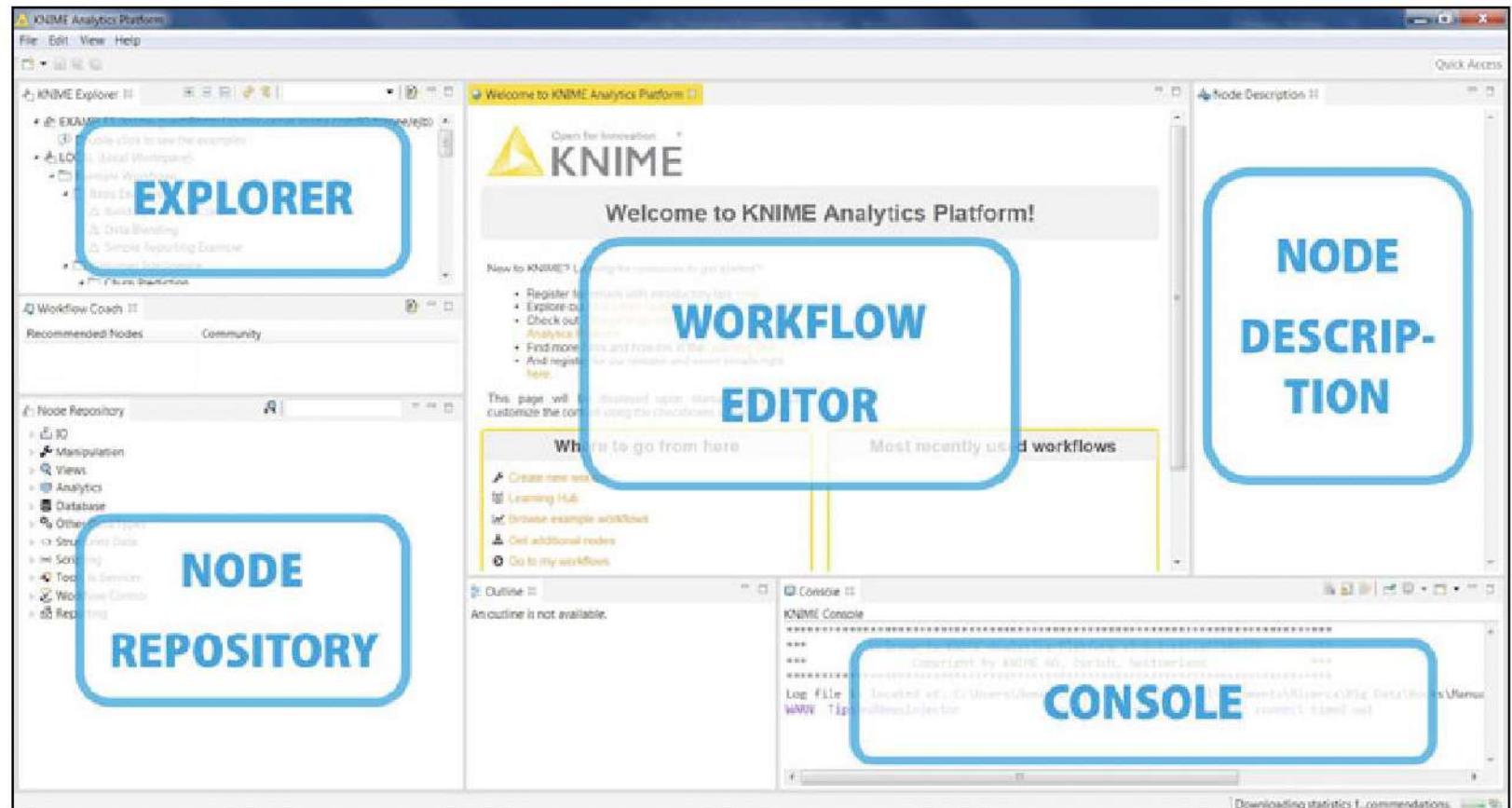
Termine	Significato
 Nodo (Node)	Unità base di computazione. Ogni nodo rappresenta un passo del processo analitico (es. leggere dati, filtrare righe, creare un modello ML, aggregare dati).
 Workflow	Insieme di nodi collegati tra loro in sequenza logica: il flusso analitico completo .

 *Analogia con Python:*

- » Nodo → funzione
- » Workflow → programma composto da funzioni collegate da variabili

L'interfaccia di KNIME: il Workbench

» La KNIME Workbench è il tuo banco di lavoro, dove esplori, costruisci e gestisci i workflow analitici.



Le cinque finestre principali

Finestra	Funzione principale
 Explorer	Gestisce e organizza i tuoi workflow (locali o remoti). Mostra la cartella LOCAL e il server EXAMPLES con modelli predefiniti.
 Node Repository	Contiene tutti i nodi disponibili, organizzati per categoria. Include un campo di ricerca per parola chiave .
 Workflow Editor	Area in cui si trascinano e collegano i nodi . Permette di aggiungere note, box e commenti. È la “tela” del tuo processo analitico.
 Node Description	Mostra la spiegazione del nodo selezionato: cosa fa, input e output richiesti.
 Console	Riporta errori, warning e log di esecuzione del workflow. Utile per il debug.

Sintesi visiva: il flusso di lavoro in KNIME

1 Node Repository → **2** Workflow Editor → **3**
Console

- » Trova i nodi necessari nella **repository**
- » Trascinali e collegali nell'**editor**
- » Esegui e controlla i risultati nella **console**
- » Esplora e riusa i workflow tramite l'**Explorer**
- » Consulta le descrizioni per capire e migliorare ogni nodo



*KNIME unisce potenza analitica e semplicità visiva:
“What You See Is What You Get”*

I nodi: unità fondamentali in KNIME

- » Ogni nodo in KNIME svolge un **compito specifico** all'interno del workflow analitico.
 - Rappresenta **un passo logico** del processo (es. lettura dati, pulizia, modellazione, esportazione).
- » **Caratteristiche principali**
 - Ogni nodo ha **un'icona grafica uniforme** (quadrato con “semaforo” sotto).
 - Funziona **solo se collegato** ad altri nodi nello stesso workflow.
 - I nodi si **eseguono da sinistra a destra**, come il flusso logico del processo.
- »  **Analogia:** I nodi in KNIME ≈ le istruzioni o funzioni in un linguaggio di programmazione.

Le connessioni tra nodi

» Il flusso del workflow

- Il workflow “scorre” da sinistra verso destra.
- Le **linee di connessione** rappresentano il passaggio dei dati tra un nodo e l’altro.

» Ogni nodo ha:

- Connettori d’ingresso (**input**) → sul lato sinistro
- Connettori d’uscita (**output**) → sul lato destro

» Tipi di connettori:

Forma / Colore

Tipo di dato trasferito

 Triangolo nero

Tabelle di numeri e testo

 Quadrato blu

Modelli, immagini, documenti, oggetti complessi



Regola d’oro: collega solo connettori compatibili (triangolo con triangolo, quadrato con quadrato).

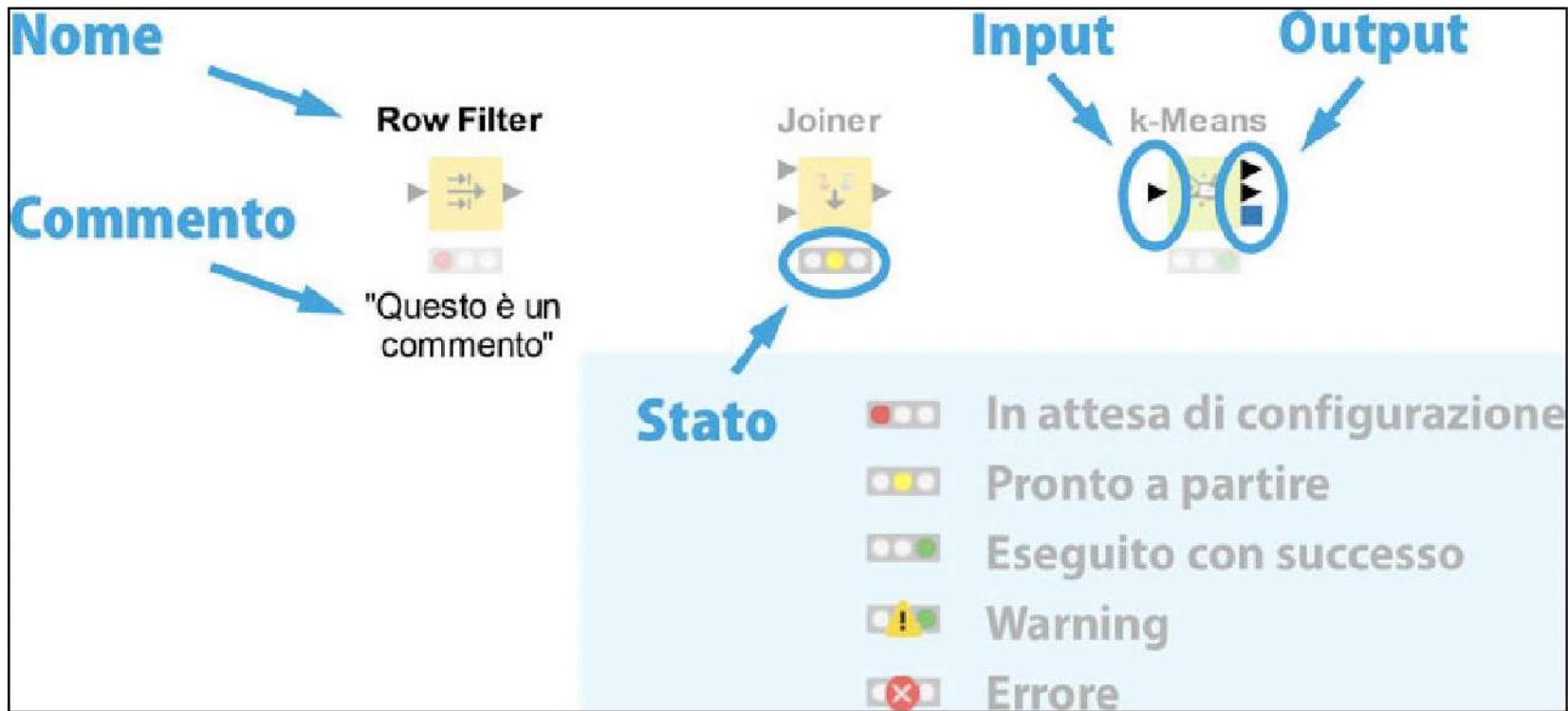
Il “semaphore” del nodo

- » Ogni nodo ha un piccolo **semaphore colorato** sotto di sé che ne indica lo stato operativo:

Colore	Significato
 Rosso	Nodo non ancora configurato o in errore
 Giallo	Nodo configurato ma non ancora eseguito
 Verde	Nodo eseguito correttamente

- »  Questo sistema visivo consente di **monitorare in tempo reale** lo stato del workflow e individuare rapidamente eventuali problemi.

Rappresentazione dei nodi in KNIME



Tipologie di nodi in KNIME

» Nodi di input:

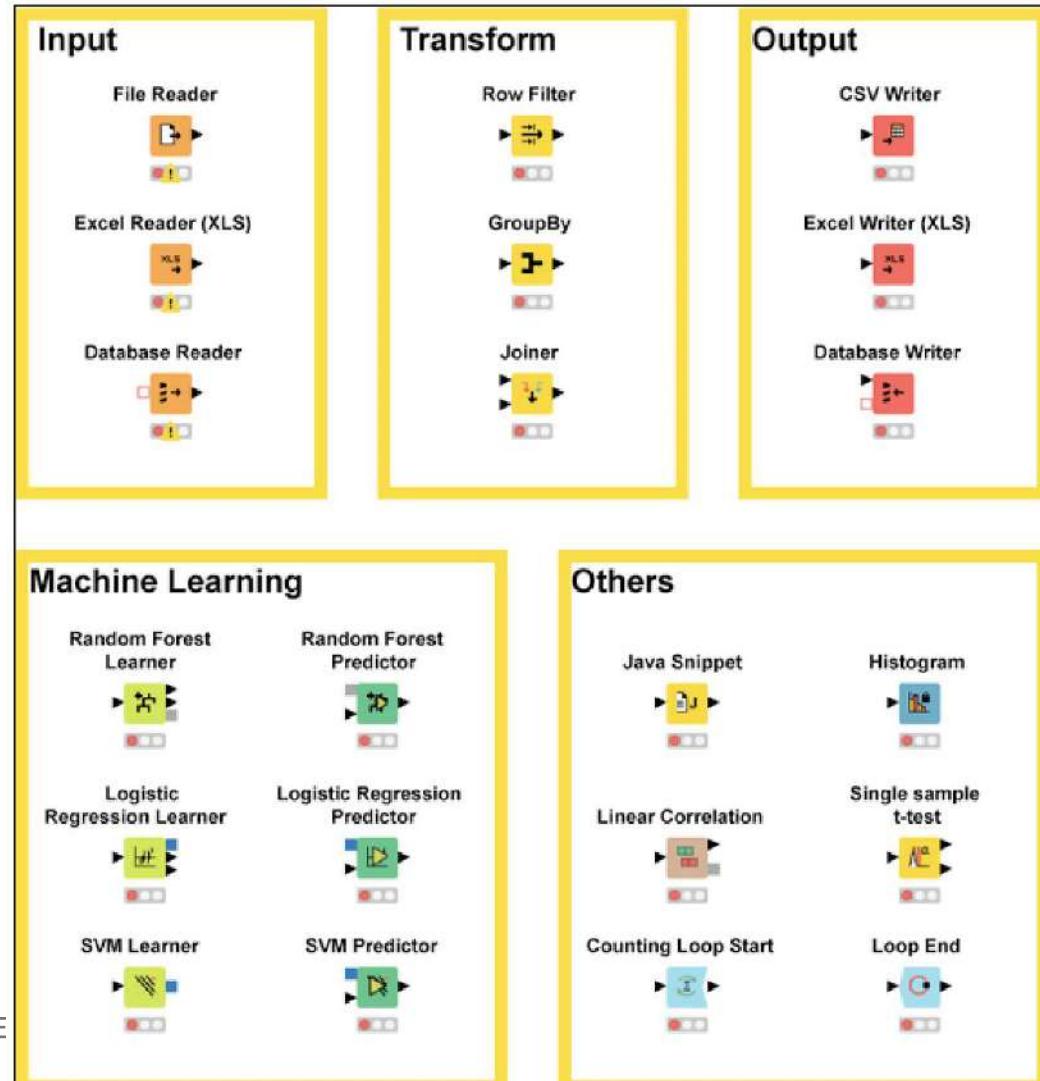
- Servono a caricare dati da file o database esterni.
- → Hanno **solo connettori in uscita (destra)**, perché forniscono dati al workflow.

» Nodi di output:

- Servono a **salvare i risultati** del workflow su file o database.
- → Hanno **solo connettori in ingresso (sinistra)**, perché ricevono dati da altri nodi.

» Nodi intermedi:

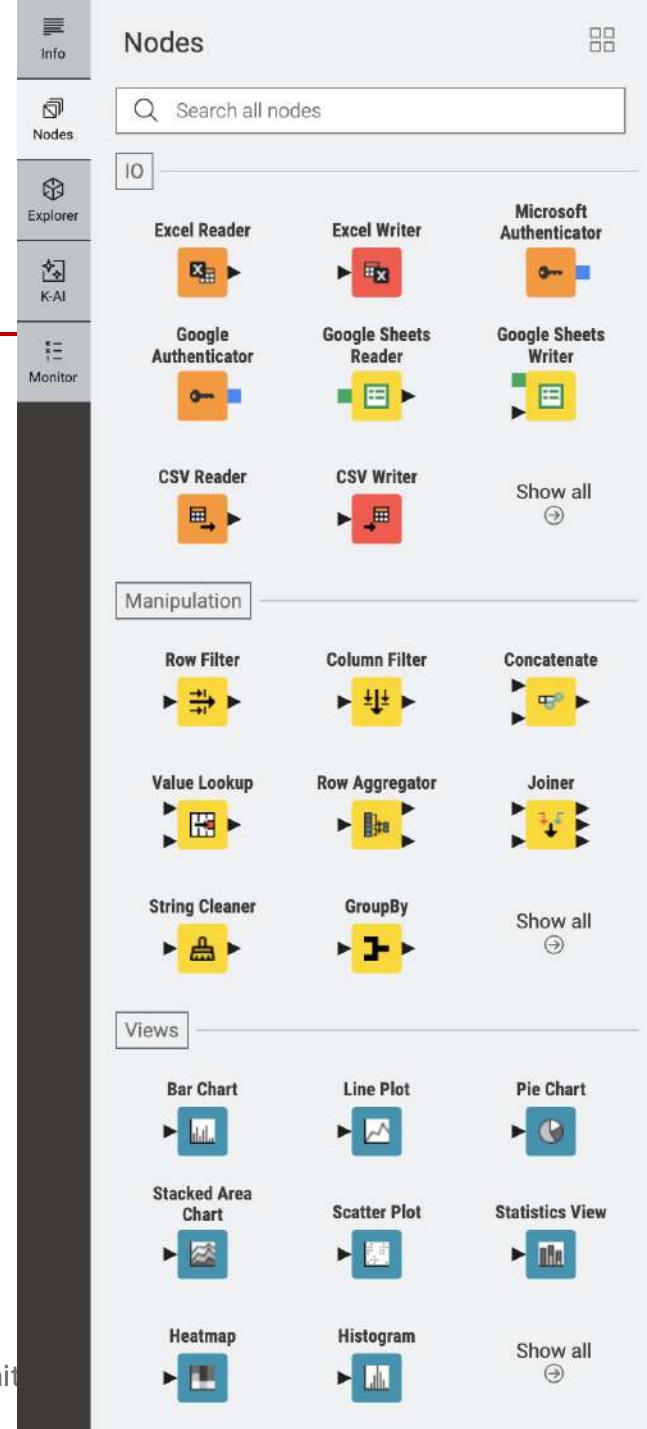
- Gestiscono le **trasformazioni e analisi** dei dati.
- → Possono avere **uno o più connettori in input e in output**, anche di tipi diversi.



Node Repository di KNIME

» Il Node Repository (Nodes) contiene i nodi che possono essere utilizzati nel flusso di lavoro

- I nodi sono organizzati in categorie
- Ogni categoria rappresenta una funzionalità specifica



Costruire il primo workflow in KNIME

» **Obiettivo:** vedere i nodi in azione e imparare a creare un workflow da zero.

» Creazione di un nuovo workflow

1. Seleziona Create a new KNIME Workflow
2. Assegna un nome (il percorso di salvataggio è scelto all'avvio dell'applicazione in “Select a directory as workspace”)
3. Clic su Create



L'editor mostrerà un workflow vuoto: una “tela bianca” su cui costruire il flusso analitico.

Home KNIME_primo_workflow +

Execute all Cancel all Reset all Local – KNIME_primo_workflow Upload

Nodes

Search all nodes

IO

Excel Reader Excel Writer Microsoft Authenticator

Google Authenticator Google Sheets Reader Google Sheets Writer

CSV Reader CSV Writer Show all

Manipulation

Row Filter Column Filter Concatenate

Value Lookup Row Aggregator Joiner

String Cleaner GroupBy Show all

Views

Bar Chart Line Plot Pie Chart

Stacked Area Chart Scatter Plot Statistics View

Heatmap Histogram Show all

Analytics

Decision Tree Predictor Decision Tree Learner Regression Predictor

Linear Regression Learner Scorer k-Means

Statistics Logistic Regression... Show all

Start building your workflow by dropping your data or nodes here.

Select a node to show its dialog.

Select a configured or executed node to show the node output.

Aggiungere un nodo al workflow

» Esempio: leggere un file Excel

1. Nella **Node Repository**, cerca:

IO → Excel Reader

2. Trascina il nodo nel Workflow Editor

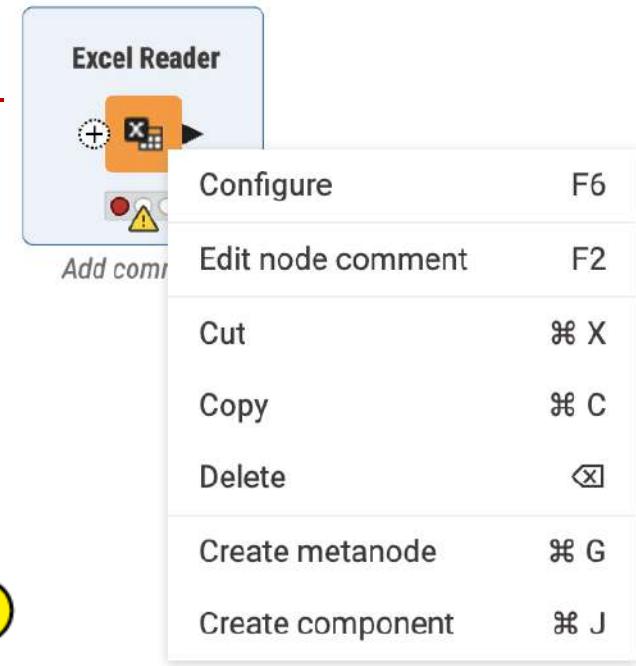
3. Il nodo apparirà con un semaforo rosso  e un warning 

→ Significa che deve essere **configurato** prima di poter funzionare.

»  *Tip:* i nodi possono essere aggiunti trascinando o tramite ricerca per parola chiave.

Configurare un nodo

- » Per configurare un nodo:
 - Clic destro sul nodo → **Configure...**
 - Si apre la finestra di configurazione:
→ Specifica il **file Excel da leggere** o altri parametri richiesti.



- »  Dopo la configurazione:
 - Il semaforo del nodo diventerà **giallo** 
 - (pronto all'esecuzione)
 - Dopo l'esecuzione corretta, diventerà **verde** 

 *In sintesi:*

- » Creare un workflow in KNIME = **trascinare, collegare, configurare e far eseguire** i nodi.

KNIME_primo_workflow

Excel Reader

This node reads Excel files (xlsx, xlsm, xlsb, and xls format). It can read a single or multiple files at the same time, however reading only one sheet per file. The supported Excel types that can be read in are string, number, boolean, date, and time but not pictures, diagrams, etc.

The data is read in and converted to the KNIME types string, Integer, long, double, boolean, local date, local time, and local date&time. Upon execution, the node will scan the input file to determine number and types of the columns and output a table with the auto-guessed structure and KNIME types.

Formulas can also be read in and reevaluated if desired. The performance of this node is limited (due to the underlying library of the Apache POI project). Reading large files takes a long time and uses a lot of memory (especially files in xlsx format when formula reevaluation is enabled).

The dialog of this node shows a preview and the file content. While the preview shows the table that will be read in having the settings applied and is updated once a setting is changed, the file content shows the content of the file as it is displayed in Excel. This allows finding row numbers and column names easier to specify different settings.

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](#).

Ports

Output ports

Type: File Table
File being read with number and types of columns guessed automatically.

File System Connection (Dynamic Import)

The file system connection.

Type: File System

File and Sheet Data Area Advanced Transformation Flow Variables Job Manager Selection Memory Policy

Input Location

Read from Local File System

Mode File Files in folder

File /Users/romina/knime-workspace/Example Workflows/Basic Examples/Combine Clean and Summarize Spreadsheet Data/Excel Reader (#6)

Select Sheet

First with data (Sheet1)
 By name Sheet1
 By position 0 (Position starts with 0.)

Preview with current settings

The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.

Row ID	Row ID	Prodotto	Quantità	Prezzo	Fattur...
Row0	Row0	Detersivo	25	4.99	124.75
Row1	Row1	Dentifricio	22	2.49	54.78
Row2	Row2	Spazzolino	14	3.00	42.00
Row3	Row3	Shampoo	25	2.75	68.75
Row4	Row4	Balsamo	13	3.12	40.56
Row5	Row5	Rasoio	11	6.49	71.39

OK Apply Cancel ?

Clic su Browse e seleziona un file Excel (preferibilmente sceglierne uno che rappresenti una tabella con il nome delle colonne indicate nella prima riga)

Excel Reader

This node reads Excel files (xlsx, xlsxm, xlsb, and xls format). It can read a single or multiple files at the same time, however reading only one sheet per file. The supported Excel types that can be read in are string, number, boolean, date, and time but not pictures, diagrams, etc.

The data is read in and converted to the KNIME types string, integer, long, double, boolean, local date, local time, and local date&time. Upon execution, the node will scan the input file to determine number and types of the columns and output a table with the auto-guessed structure and KNIME types.

Formulas can also be read in and reevaluated if desired. The performance of this node is limited (due to the underlying library of the Apache POI project).

Reading large files takes a long time and uses a lot of memory (especially files in xlsx format when formula reevaluation is enabled).

The dialog of this node shows a *preview* and the *file content*. While the preview shows the table that will be read in having the settings applied and is updated once a setting is changed, the file content shows the content of the file as it is displayed in Excel. This allows finding row numbers and column names easier to specify different settings.

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](#).

Ports **Options** **Views**

Output ports

► **Type:** File Table
File being read with number and types of columns guessed automatically.

File System Connection (Dynamic Import)
The file system connection.

Local – KNIME_primo_workflow

Excel Reader

In Data Area, è possibile selezionare Column Names spuntando la casella Use value in row ...

Dialog - 3:1 - Excel Reader

Data Area (tab circled in blue)

Use values in row 1

Read Area

Whole sheet

Range by row and column: Column A to and row 1 to .

Column Names

Use values in row 1

Skip

Empty rows Empty columns Hidden rows Hidden columns

Preview **File Content**

Preview with current settings

The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.

Row ID	Prodotto	Quantità	Prezzo	Fattur...
Row0	Detersivo	25	4.99	124.75
Row1	Dentifricio	22	2.49	54.78
Row2	Spazzolino	14	3.00	42.00
Row3	Shampoo	25	2.75	68.75
Row4	Balsamo	13	3.12	40.56
Row5	Rasoio	11	6.49	71.39

OK Apply Cancel

Click su Apply e poi Ok, il semaforo diventerà giallo

Excel Reader



Esecuzione e controllo dei nodi in KNIME

▶ Esecuzione e controllo dei nodi in KNIME

» Una volta configurato, il nodo mostra il **semaforo giallo**  : è pronto per l'esecuzione.

» Per eseguirlo:

- clic destro → **Execute**, oppure
- seleziona il nodo e premi **F7**.

→ Il nodo legge il file Excel e **rende disponibile la tabella** ai nodi successivi nel workflow.

» KNIME consente di **monitorare passo per passo** l'esecuzione del workflow e di visualizzare i risultati intermedi.

» Per vedere l'**output di un nodo**, ad esempio dell'*Excel Reader*,:

- clic destro → **Output table**, oppure
- premi **Shift + F6**.

» Nella finestra dei risultati ogni colonna mostra:

- **Nome**,
- **Icona del tipo di dato** (S = stringa, I = intero, D = decimale, ecc.).

» Il pannello **Spec – Columns** consente di esplorare le specifiche tecniche (tipo di dato, valori min/max, ecc.).

 *In sintesi:* dopo l'esecuzione, puoi sempre **ispezionare i dati generati** da ogni nodo per controllare e comprendere il flusso del workflow.

Clic sulla voce Execute del menu contestuale o sul nodo.

Visualizzeremo la tabella letta dal file Excel



KNIME_primo_workflow

Home Preferences Menu Sign in

Execute Cancel Reset

Local – KNIME_primo_workflow

Upload

Info Nodes Explorer K-AI Monitor

Excel Reader

This node reads Excel files (xlsx, xlsm, xlsb, and xls format). It can read a single or multiple files at the same time, however reading only one sheet per file. The supported Excel types that can be read in are string, number, boolean, date, and time but not pictures, diagrams, etc.

The data is read in and converted to the KNIME types string, integer, long, double, boolean, local date, local time, and local date&time. Upon execution, the node will scan the input file to determine number and types of the columns and output a table with the auto-guessed structure and KNIME types.

Formulas can also be read in and reevaluated if desired. The performance of this node is limited (due to the underlying library of the Apache POI project).

Reading large files takes a long time and uses a lot of memory (especially files in xlsx format when formula reevaluation is enabled).

The dialog of this node shows a preview and the file content. While the preview shows the table that will be read in having the settings applied and is updated once a setting is changed, the file content shows the content of the file as it is displayed in Excel. This allows finding row numbers and column names easier to specify different settings.

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.

Ports Options Views

Output ports

Add comment

100% +

1: File Table Flow Variables

Rows: 6 | Columns: 4

Table Statistics

Fatturato

Prezzo

Quantità

Prodotto

RowID

#

Detersivo

1 Detergente

Dentifricio

2 Dentifricio

Spazzolino

3 Spazzolino

Shampoo

4 Shampoo

Balsamo

5 Balsamo

Rasoio

6 Rasoio

#	RowID	Prodotto	Quantità	Prezzo	Fatturato
1	Row0	Detersivo	25	4.99	124.75
2	Row1	Dentifricio	22	2.49	54.78
3	Row2	Spazzolino	14	3.00	42.00
4	Row3	Shampoo	25	2.75	68.75
5	Row4	Balsamo	13	3.12	40.56
6	Row5	Rasoio	11	6.49	71.39

Nota – Tipi di dato e buone pratiche in KNIME



Tipi di dato principali

Prodotto	String
Quantità	Number (Integer)
Prezzo	String
Fatturato	String

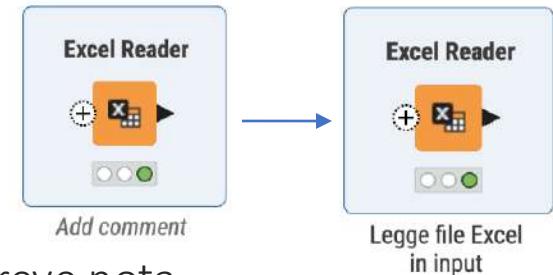
Ogni colonna in KNIME ha un **data type** specifico. I più comuni sono:

- » **Integer number** → numeri interi (32 bit)
- » **Long number** → interi lunghi (64 bit)
- » **Double number** → numeri decimali
- » **String** → testo
- » **Boolean** → valori logici (0 o 1)
- » *(Esistono anche collezioni, come liste e insiemi...)*



Buona pratica: commentare i nodi

- » Ogni nodo può (e dovrebbe) essere **descritto** con una breve nota.
- » Fare **doppio clic sul testo sotto il nodo** per rinominarlo.
- » I commenti rendono i workflow **più leggibili, riutilizzabili e modificabili** da altri utenti (e da sé stessi in futuro).

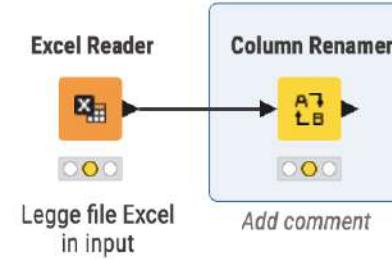
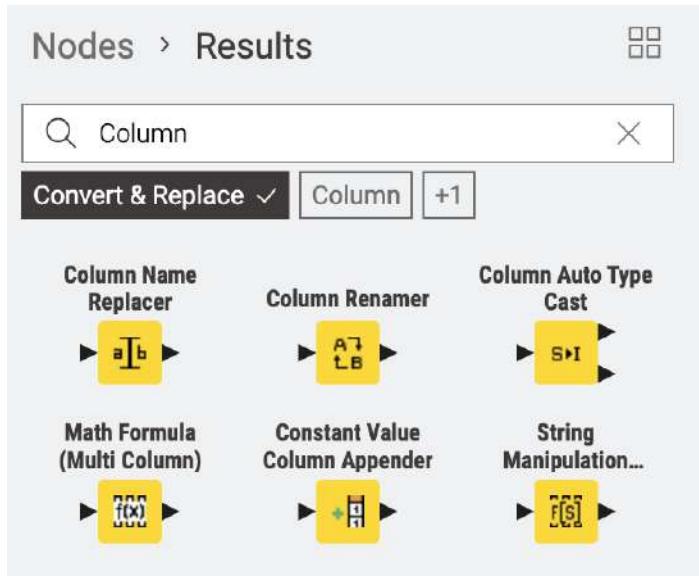


Aggiungere e collegare un nuovo nodo

1. Cerca, nel Node Repository, Column Rename in

Manipulation → Column → Convert & Replace → Column Renamer

2. Trascinalo **a destra** del nodo *Excel Reader*
 3. Collega il **triangolo nero** di uscita del primo nodo con il **triangolo nero** di **ingresso** del secondo
- L'output del primo nodo diventa l'input del secondo
4. Seleziona il nodo e imposta i parametri (finestra sulla destra)



Excel Reader → **Column Renamer**
 Input tabella prodotti Add comment

Legge file Excel in input

Column Renamer

Column	New name
Fatturato	Revenue
Prezzo	Price
Quantità	Quantity
Prodotto	Article

+ Add column

Discard Execute Apply

Indicare il nuovo nome per ogni Colonna che si vuole rinominare

Completere il primo workflow in KNIME

» Aggiungi un nodo di **output** per salvare i risultati:

IO → CSV Writer

» Collega il **CSV Writer** al nodo *Column Renamer*

» Nella configurazione:

- Specifica **nome** e percorso del file CSV da creare.
- Attiva l'opzione **Write column header** per mantenere le intestazioni

» Esegui il nodo o l'intero workflow

→ Il file CSV generato conterrà le colonne rinominate in inglese

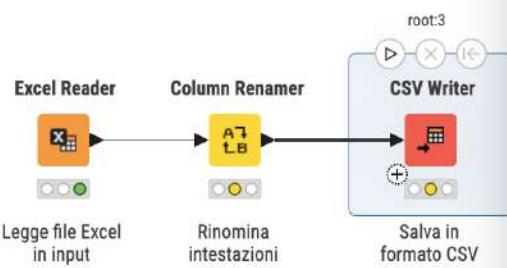
Settings

Advanced Settings

Comment Header

Encoding

Flow Variables



Output location

Write to

Local File System

File

/Users/romina/knime-workspace/Example Workflows/Basic Examples/output_prodotti.csv

Write options

 Create missing foldersIf exists: overwrite append fail

Format

, Column Delimiter

System Default

Row Delimiter

" Quote Char

" Quote Escape Char

Header

 Write column header

K - Execute

Apply

Cancel



output_prodotti

	Article	Quantity	Price	Revenue
1	Detersivo	25	4.99	124.75
2	Dentifricio	22	2.49	54.78
3	Spazzolino	14	3.00	42.00
4	Shampoo	25	2.75	68.75
5	Balsamo	13	3.12	40.56
6	Rasoio	11	6.49	71.39
7				

File CSV creato

Flow Variables

Count: 1