

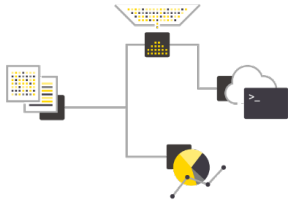
Highlights and Updates

KNIME Software

April 18, 2023



~~Intermezzo:~~ KNIME Software



KNIME Analytics Platform

 FREE, OPEN SOURCE

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 Business Hub

 COMMERCIAL

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.

Create

Blend & Transform



Access, merge, and transform all of your data

Model & Visualize



Make sense of your data with the tools you choose

Productionize

Deploy & Manage



Support enterprise-wide data science practices

Consume & Interact



Leverage insights gained from your data

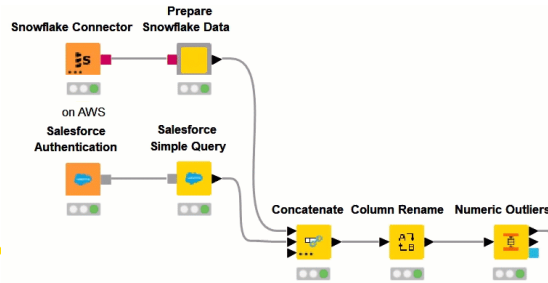
A Complete Analytics Platform for All Data Workers



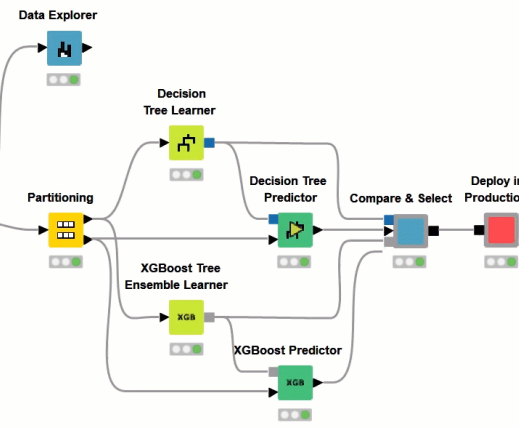
300+ Data Sources

- Files, Docs
- Audio, Images
- Databases
- Apps
- Cloud
- SAP
- APIs
- Sensor data

Blend & Transform



Model & Visualize



Deploy & Manage

- REST
- Edge
- Data Apps
- BI Tools
- Docs & Emails

Advanced Methods

Scripting Languages

AI/ML Libraries

Distributed & Cloud Execution

KNIME Analytics Platform Version 5

The screenshot displays the KNIME Analytics Platform interface for a workflow titled "Building a Simple Classifier". The workflow consists of the following nodes:

- CSV Reader**: Reading adult.csv
- Partitioning**: Random drawing 80% upper port 20% lower port
- Decision Tree Learner**: Train to predict class 'income'
- Decision Tree Predictor**: Apply decision tree model to test set
- Scorer**: Node 12
- Scatter Plot**: Age vs. number-hours

The left sidebar shows the "Partitioning" node configuration. The input table is split into two partitions (i.e. row-wise), e.g. train and test data. The following options are available in the dialog:

- Input ports**:
 - Type: Table to partition
 - Table to partition.
- Output ports**:
 - Type: First partition (as defined in dialog)
 - First partition (as defined in dialog).
 - Type: Second partition (remaining rows)
 - Second partition (remaining rows).

A search for compatible nodes is also visible, showing the following results:

- Table Splitter
- Table Cropper
- Column Filter
- Row Filter
- Joiner
- Row to Column Names
- Concatenate
- Missing Value
- String Manipulation
- Number To String
- String to Date&Time
- Rule-based Row Filter


The data table below shows the following columns: #, RowID, age, workclass, fnh, al-st..., occupation, relations..., race.

#	RowID	age	workclass	fnh	al-st...	occupation	relations...	race		
1	Row0	39	State-gov	77		Adm-clerical	Not-in-family	White		
2	Row2	38	Private	21		Handlers-clean...	Not-in-family	White		
3	Row3	53	Private	23		Handlers-clean...	Husband	Black		
4	Row4	28	Private	33		Prof-specialty	Wife	Black		
5	Row6	49	Private	16		Other-service	Not-in-family	Black		
6	Row8	31	Private	45		Prof-specialty	Not-in-family	White		
7	Row9	42	Private	159449	Bachelors	13	Married-civ-spo...	Exec-managerial	Husband	White
8	Row11	30	State-gov	141297	Bachelors	13	Married-civ-spo...	Prof-specialty	Husband	Asian-Pac-Islan...
9	Row12	29	Private	12222	Bachelors	13	Never-married	Adm-clerical	Own child	White

Hitchhiker's Guide to KNIME Analytics Platform Version 5

Stage

Apr 18 04:45PM–05:30PM MESZ

 Add to calendar

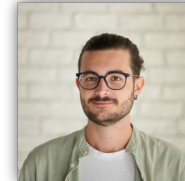
Workshop: Hitchhiker's Guide to KNIME Analytics Platform Version 5

Elisabeth Richter, Emilio Silvestri, Schalk Gerber - KNIME

On March 1st, the early access KNIME Analytics Platform Version 5 was released. With that, we are providing an improved UX/UI that makes it easier for new users to get started by having a more intuitive workflow editor, an enhanced node search, and a new set of data wrangling functionalities.

In addition, KNIME Analytics Platform Version 5 provides the “Starter Perspective”, a curated set of nodes, that ensures a friendly entry point to advanced analytics and an easy transition from spreadsheets to visual workflows.

Join this workshop to get familiar with the improved UX/UI and learn how to find your way around the KNIME workbench. Build your first workflow and explore the “Starter Perspective” and its new spreadsheet manipulation and data wrangling features.



Requirements of a Complete Data Science Platform

KNIME Analytics Platform Version 5

© 2023 KNIME AG. All rights reserved. 5

New Extensions

Scikit-learn (sklearn)

- The machine learning toolkit in Python
- Algorithms can now easily be made available in KNIME
- Will be a community effort

Geospatial Analysis

Center for Geographic Analysis at Harvard University

[\(Getting started with Geospatial Analysis in KNIME\)](#)



© 2023 KNIME AG. All rights reserved.

15

KNIME

Broad

- Ease of use & onboarding
- Support for common use cases
- Extensible & integrative

Specific

- Advanced capabilities
- Domain-relevant
- Legacy tool compatible

Database Framework Enhancements

- KNIME's database nodes allow connecting to databases and transfer and manipulate tables

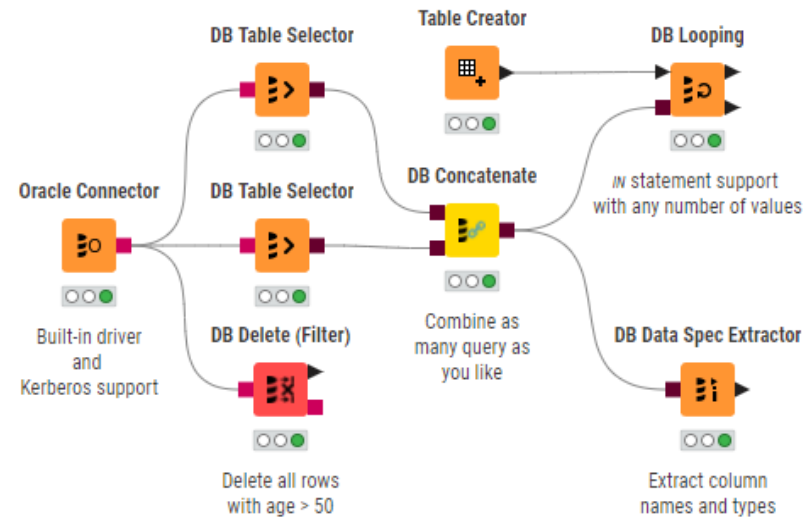
- New nodes:

- DB Row Manipulator
- DB Looping (e.g. IN queries)
- DB Concatenate
- DB Delete (Filter)
- DB Data Spec Extractor

- Improved Oracle connector

- Kerberos authentication support

- New Microsoft Azure services connection



Snowflake Machine Learning Model Push-Down

1. Selection.

Selects the "calls_contracts" table to use for prediction.

Snowflake Connector



Change connection information to your Snowflake account

DB Table Selector



Select existing table with new data

2. Deployment and Execution.

Convert the customer churn prediction model into a User-Defined Function in Snowflake.

H2O MOJO Reader



Read the learned model

Snowflake H2O MOJO Predictor (Classification)



Apply H2O model at scale in Snowflake

3. Persisting.

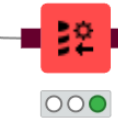
The prediction result can either be read back into KNIME or persisted as new database table within Snowflake.

DB Reader



Read data into KNIME

DB Connection Table Writer



Persist prediction result in Snowflake

<https://www.knime.com/blog/churn-predictor-knime-snowflake>

And there's more...

- *Call Workflow* nodes have been streamlined
- New *Raw Container Input* and *Output* can be used to build a generic REST API
- *Column Expressions* now support multi-row formula
- Native Apple Silicon build of KNIME (since AP 4.7)

KNIME Extensions



Why build a KNIME Extension?

Data Engineering Team

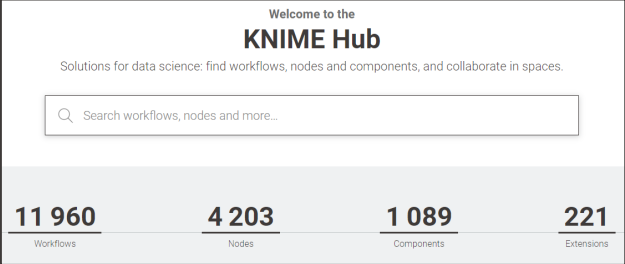


Pure-Python
Nodes



KNIME

Finance Team



Python Integration on Steroids

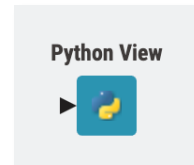
- Scripting in KNIME



- Python integration performance boost
 - No latency: Zero-copy data transfer between KNIME and Python
 - Work with KNIME tables larger than main memory

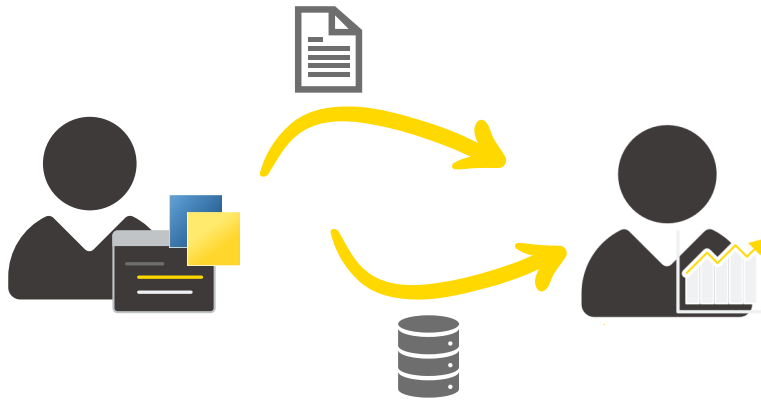
- Python View node

- Interactive views



KNIME Extensions in Python

- Pythonic API to interact with KNIME
 - Report progress, access flow variables, use logging, set warnings, ...



```
import knime.extension as knext
import pandas as pd

@knext.node(name="Geo Distances", node_type=knext.NodeType.MANIPULATOR,
            icon_path="...", category="/")

@knext.input_table(name="Distances Table", description="Location pairs")
@knext.input_table(name="Coordinates Table", description="Lat/Long...")

@knext.output_table(name="Output Data", description="Distances...")
@knext.output_view("Scatter View", "Distances in a scatter plot")

class GeoDistances:
    """
    This node is able to compute distances between locations
    given latitude and longitude coordinates
    """

    location_column = knext.ColumnParameter("Location Column", "...", port_index=1)
    lat_column = knext.ColumnParameter("Latitude Column", "...", port_index=1)
    long_column = knext.ColumnParameter("Longitude Column", "...", port_index=1)
    output_column_name = knext.StringParameter("Output Column name", "...", "Distance")

    def configure(self, configure_context, input_schema_1, input_schema_2):
        """
        # Return output table schema
        return input_schema_1.append(
            knext.Column(knext.double(), name=self.output_column_name)
        )

    def execute(self, exec_context, input_1, input_2):
        """
        # TODO: do something useful here
        trips_df = pd.DataFrame()

        # Return results
        return knext.Table.from_pandas(trips_df), knext.view_matplotlib()
```

New Extensions

Geospatial Analysis



Center for
Geographic Analysis
at Harvard University



Spatial IO



Spatial Manipulation



Open Datasets



Spatial Transformation



Spatial Conversion

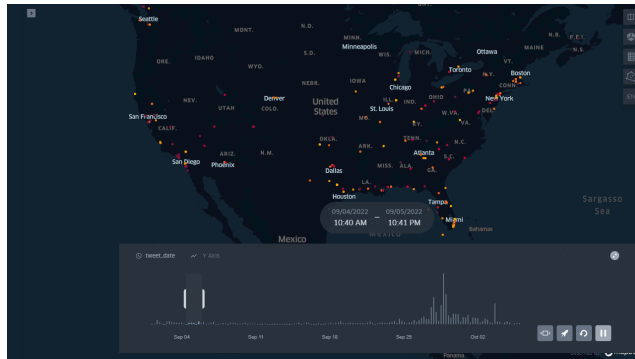


Spatial Visualization



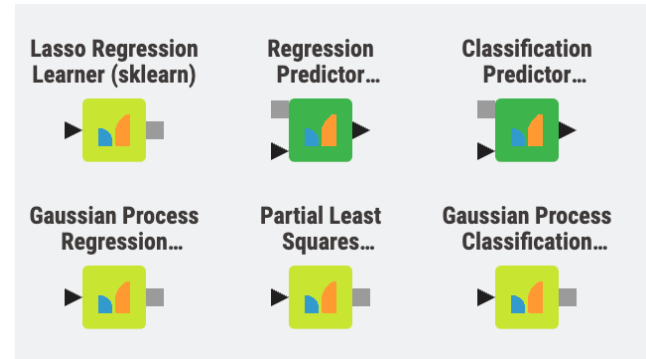
Spatial Calculation

[\(Getting started with Geospatial Analysis in KNIME\)](#)



Scikit-learn (sklearn)

- *The* machine learning toolkit in Python
- Algorithms can now easily be made available in KNIME
- Will be a community effort



Demo

CO₂ Forecasting




New Extensions

Redfield NLP Nodes


v 1.1.1

Included nodes




Spacy Lemmatizer

Community Nodes | Redfield NLP



Spacy Model Selector

Community Nodes | Redfield NLP




Spacy Morphologizer

Community Nodes | Redfield NLP

Redfield BERT Nodes


v 1.0.1

Included nodes




BERT Classification Learner

Community Nodes | Redfield BERT Nodes



BERT Embedder

Community Nodes | Redfield BERT Nodes



BERT Model Selector

Community Nodes | Redfield BERT Nodes

Trains a text classification model on top of provided BERT model. The model is external.

Maps String or Document columns to a vector using the provided BERT model. The model is external.

The node allows to select a model from Text Hub or HuggingFace. The trusted model is external.

Apprise Nodes


extension allows notifications through Apprise

Included nodes

another rather simple Node in the collection of Apprise Nodes. It can be used to...

the most simple Node in the collection of the Apprise Nodes. It can be used to s...

node can be used to send out a series of notifications defined in a table. It expect...




Workflow Diagram

```

graph LR
    A[Test Plan Information] --> B[Column Appender]
    B --> C[Create EC2 Instance]
    C --> D[Describe EC2 Instance]
    D --> E[Run Shell Command on EC2 Instance]
    E --> F[Manage EC2 Instance]
    
```

© 2023 KNIME AG. All rights reserved.

17



BYOE: Build Your Own Extension

KNIME Documentation

Read or download documentation for KNIME Software

Create a New Java based KNIME Extension

Create a New Python based KNIME Extension

Stage

Apr 19 10:45AM–11:30AM CEST

 Add to calendar

Workshop: Maximizing the Potential of Data Science with KNIME and Python

Paolo Tamagnini, Mahantesh Pattadkal, Carsten Haubold - KNIME

In large organizations, higher productivity results from collaboration on preferred tools. Regardless of

preferences, KNIME offers a solution for fast and productive collaboration! This

new Python can be used in conjunction with KNIME.

low-code approach. You will learn how to wrap Python scripts in KNIME

and part of the webinar, we will go a pure-code approach and show Python

KNIME nodes and extensions in Python. This new capability offers data

framework to integrate existing and custom Python libraries into KNIME by

configuration, execution, as well as dialog definition via a Pythonic API.

Networking

Apr 18 03:30PM–04:15PM CEST

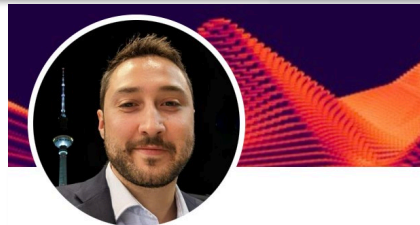
 Add to calendar

Networking Session, Partner Exhibition & Virtual Breakouts & Birds of a Feather Tables

Visit the onsite KNIME Booths for Demo, Press and Learning.



Steffen Fissler · 1st
Community Extensions Manager – KNIME



Stephen Rauner · 2nd
Partner Technology Manager – KNIME

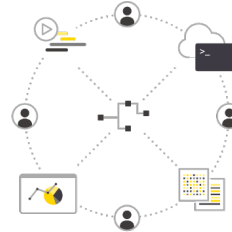
~~Intermezzo:~~ KNIME Software



KNIME Analytics Platform

FREE, OPEN SOURCE

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 Business Hub

COMMERCIAL

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.

Create

Blend & Transform



Access, merge, and transform all of your data

Model & Visualize



Make sense of your data with the tools you choose

Productionize

Deploy & Manage



Support enterprise-wide data science practices

Consume & Interact



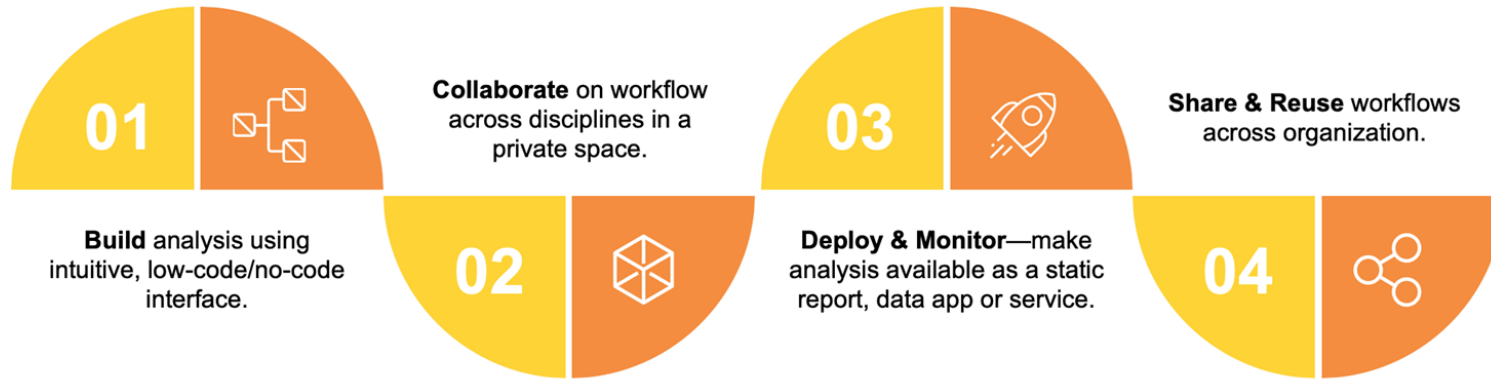
Leverage insights gained from your data

KNIME Business Hub

Megan Kattawar



A Complete Platform: From Ideation to Production



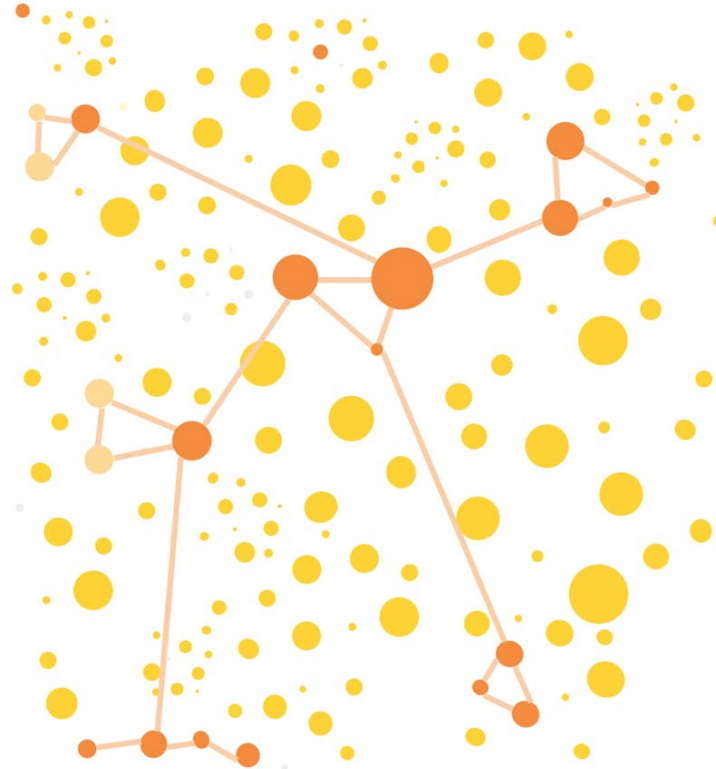
KNIME Analytics
Platform

KNIME Community Hub

KNIME Business Hub

A Single Platform for All Data Users

- **Data Experts (users)**
 - No-code & low-code workflows
 - Create custom nodes for sharing & re-use
 - Extensions for advanced analysis
- **Business & Domain Experts (users)**
 - No-code & low-code workflows
 - Library of pre-built workflows for fast start
 - Fast & instant access to insights, without IT
- **End Users (non-users)**
 - Data apps for insight insights
 - APIs, running behind-the-scenes
 - Automated reports



MLOps & IT

- Secure deployment
- Centralized governance

Scaling Data Science Impact

Ease of Onboarding

Get started quickly with a community-built library of solutions.



Ease of Collaboration

Work between disciplines & expert levels. Share & reuse expertise.



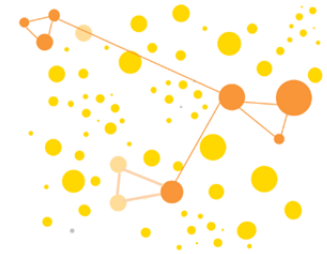
Ease of Deployment

Enables data experts to easily deploy data apps and services to end users.



Built for Scale

Cloud-native, no limit to users. Easy to administrate & maintain.



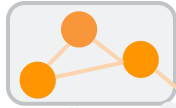
Open-source approach ensures frictionless adoption, low barrier to entry and low TCO

Demonstration

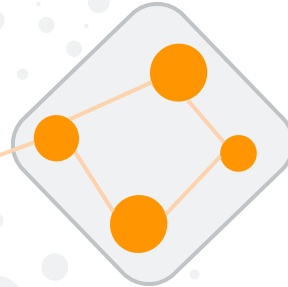
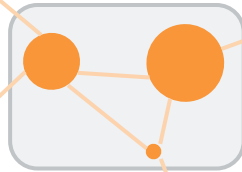


IT Sets Guardrails, Teams Self-Manage

CoE Team



Marketing Team



R&D Team



Built to Scale

IT controls:

- Allocation of cores
- Seats per team
- Execution configuration (optionally)

Teams control:

- Users
- Execution context
- Execution configuration
- Monitoring

De-Centralized Administration

KNIME Business Hub in the Organization

IT only allocates seats & execution cores per team.

Seats allocated by teams

Marketing Team



R&D Team



CoE Team



Read, write or execute permissions determined by team

Marketing spaces



Knowledge sharing



Development



Knowledge sharing



Development



Deployment & monitoring by team

Scheduled execution for Monthly Reporting



REST API



Data App



Data App for Leadership



Data App for Marketing



REST API for Production



Execution context, determined by team

Marketing execution context
(8 cores, no GPU)

KNIME expert execution context
(8 cores, GPU)

Dev & Test execution context
(4 cores, GPU)

Prod execution context
(8 cores, GPU)

KNIME Community Hub - Teams Edition

Pricing for KNIME Community Hub

EUR USD

Personal plan

€ 0 / month

Sign up for free

What's included

- ✓ Integration with open-source [KNIME Analytics Platform](#)
- ✓ Private spaces for self-use
- ✓ Collaboration in public spaces

[Discover all features](#)

Team plan


starts from € 250 / month

Get started

What's included

- ✓ Everything from Personal plan; plus
- ✓ Collaboration with teams in private spaces
- ✓ Extend disk storage
- ✓ Centralized billing

[Discover all features](#)

 **Free for teaching and non-profit research**

[Join KNIME Educators Alliance](#) to learn more



Thank You.