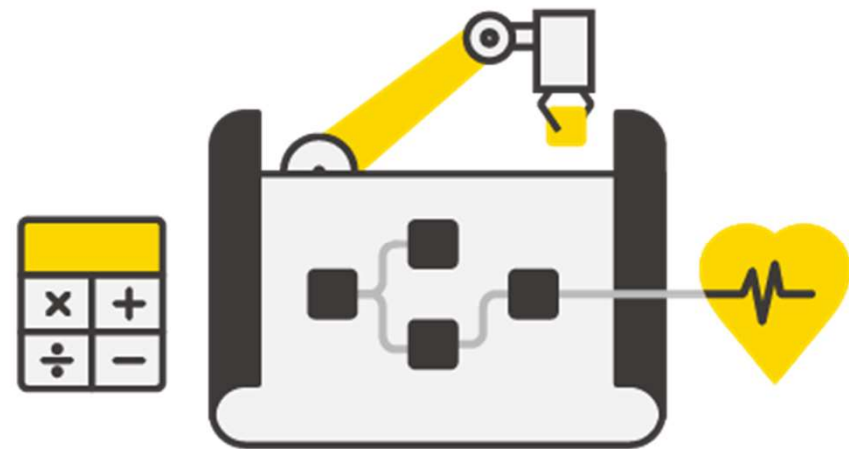




HUPDATA
Data Analysis Solutions

KNIME Data Talks – Brazil

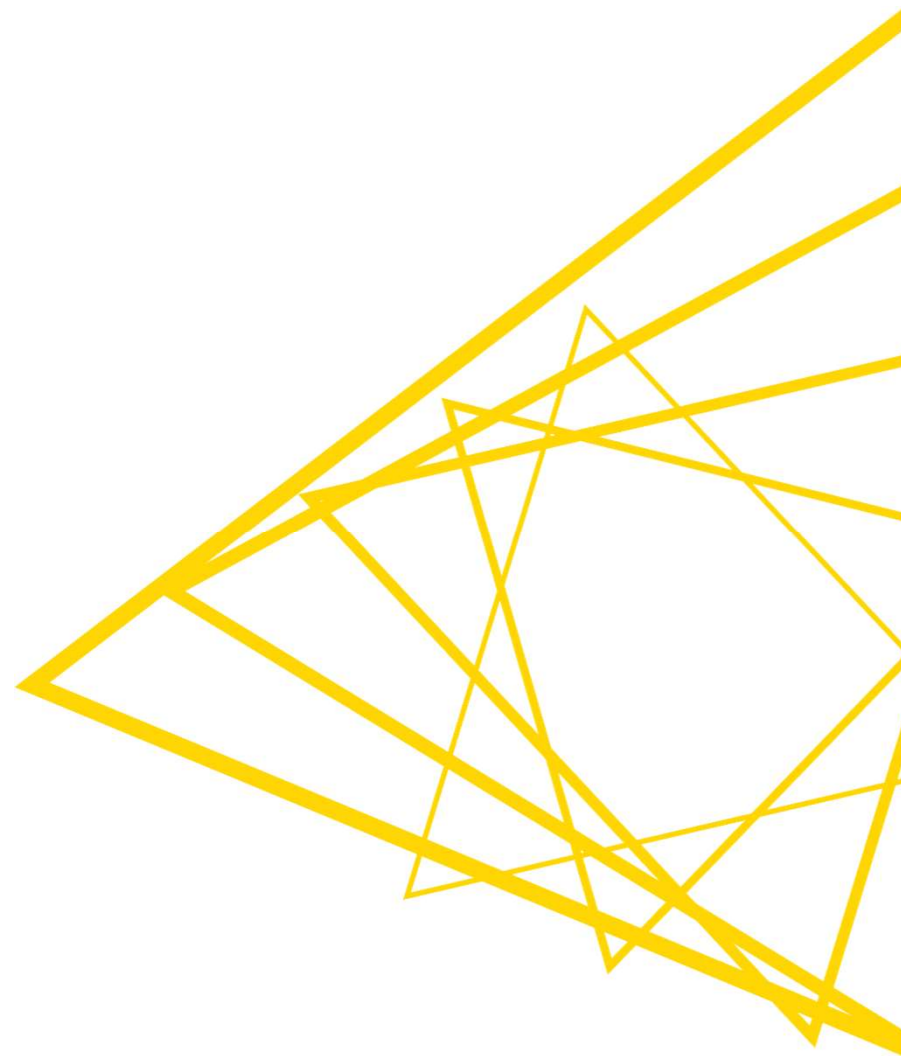
Novidades KNIME Analytics Platform
Tiago Silva



May 18-21, 2021

Agenda

- Novas funções para usuários KNIME
- Facilidades com serviços de Nuvem
- Novas integrações Python
- Novas funções no KNIME Hub

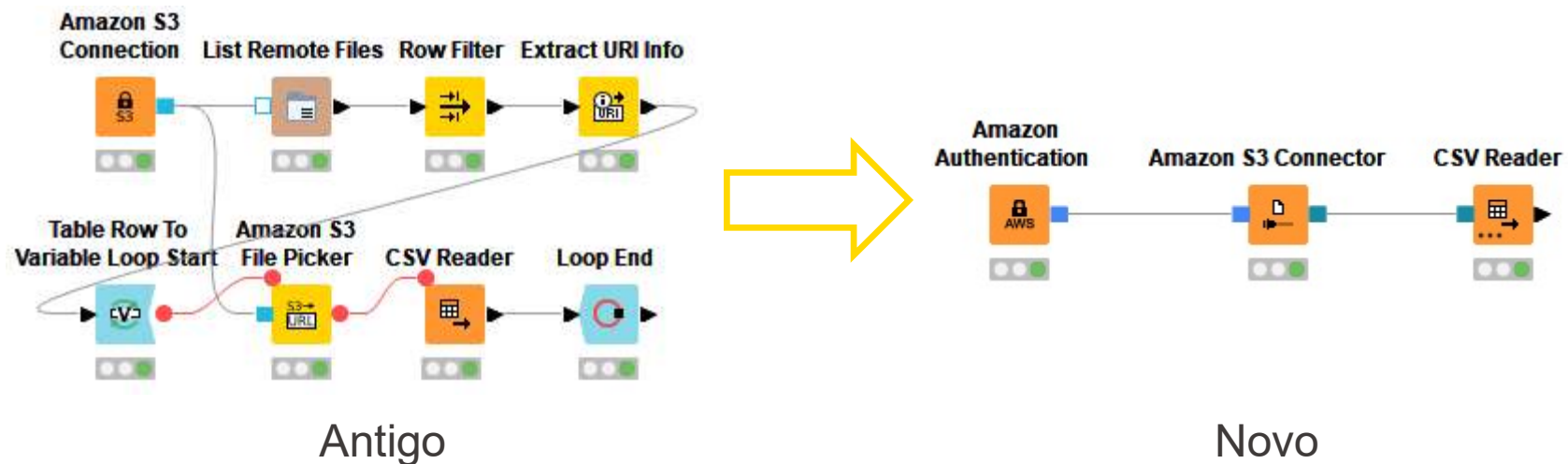


Estrutura de manipulação de arquivos– out of Labs



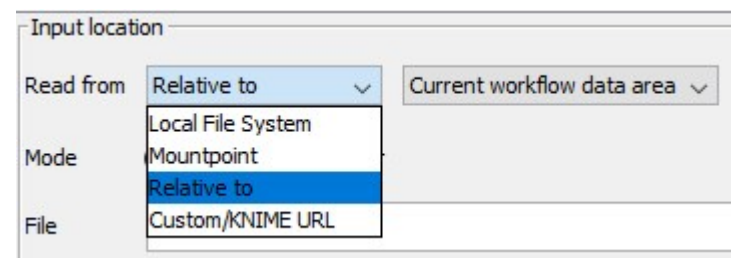
Estrutura de manipulação de arquivos - Benefícios

- Experiência de usuário consistente em todos os nodes e sistemas de arquivos
- Fácil de migrar fluxos de trabalho de um sistema de arquivos para outro
- Estrutura poderosa que nos permite integrar mais e mais sistemas de arquivos
- Melhorias de desempenho

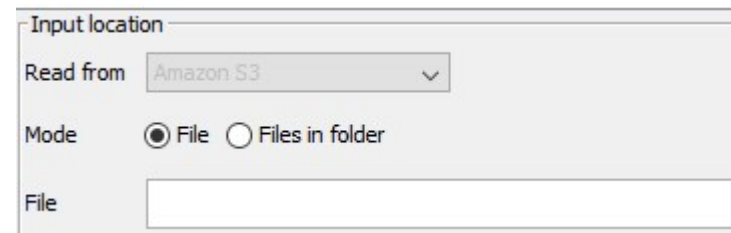


File Systems – Sistema de arquivos

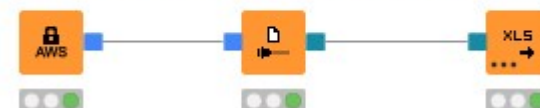
- Arquivos de conexão convencionais
 - Arquivo local
 - Ponto de montagem
 - Relative to (workflow, mountpoint, workflow data area)
 - Custom/KNIME URL
- Conexão com ambientes:
 - Microsoft Azure
 - Google
 - Amazon
 - Databricks
 - BigData file systems (HDFS, HTTPFS, ...)
 - On-premise (e.g., SSH, FTP, ...)
- Nodes de manipulação com portas dinâmicas



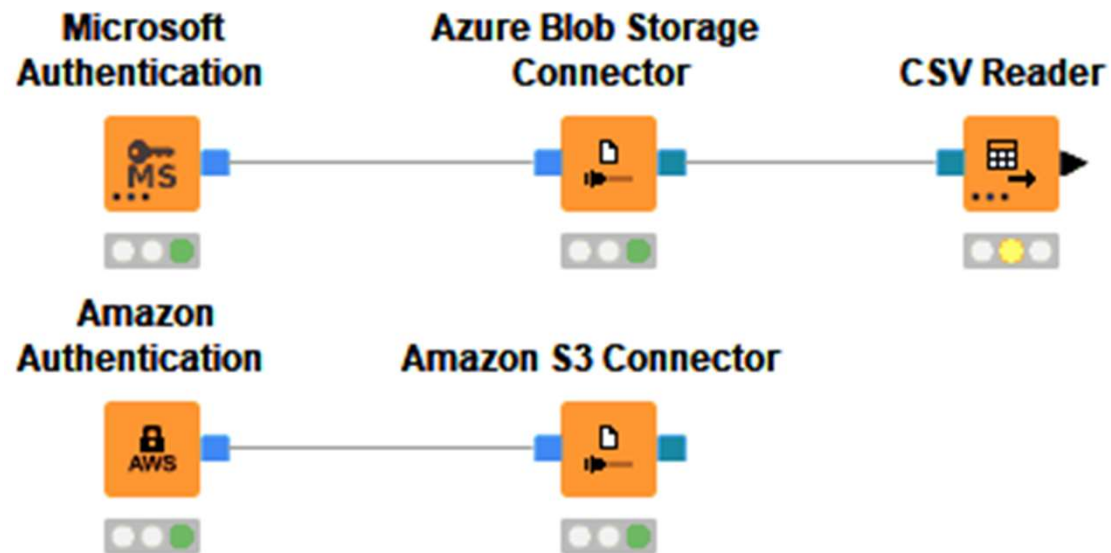
Excel Reader (XLS)



Amazon Authentication Amazon S3 Connector Excel Reader (XLS)



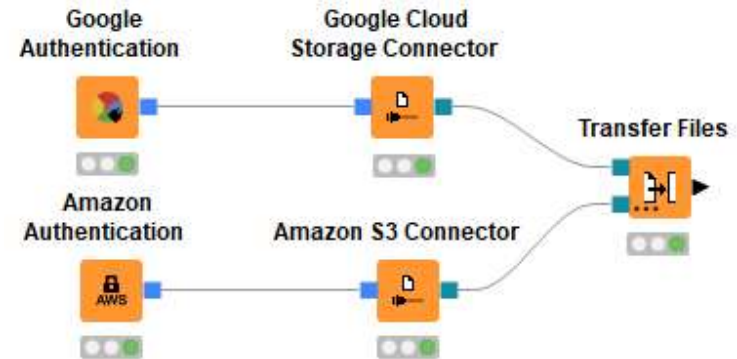
Sistema de arquivos agnóstico



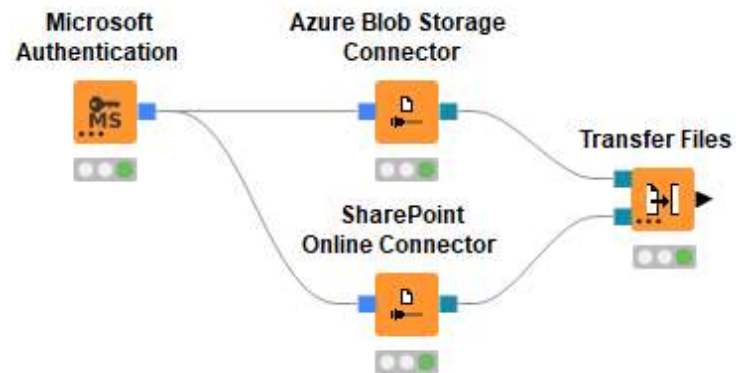
Total Flexibilidade



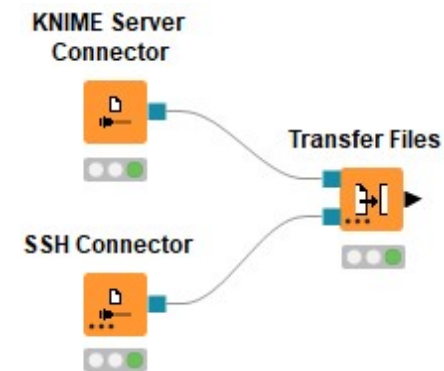
Arquivos locais e ponto de montagem



Cruzamento entre nuvens

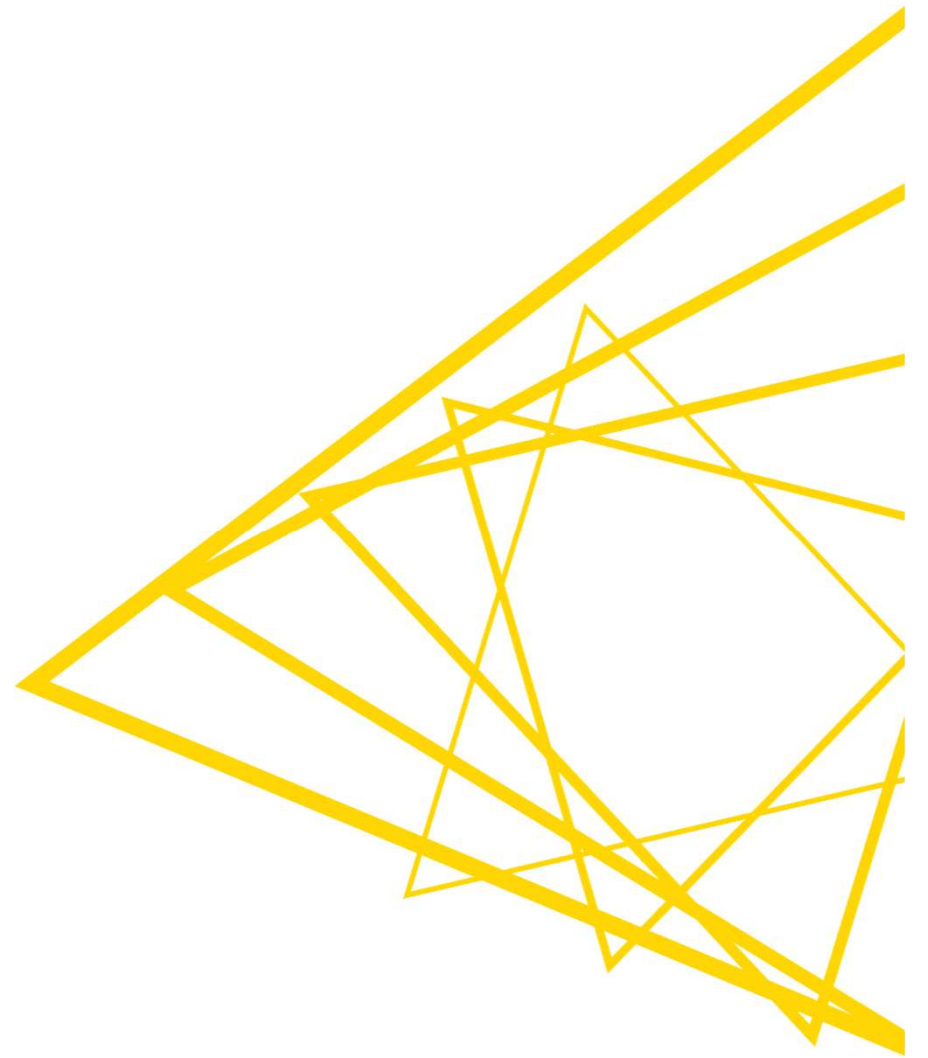


Mesmo ambiente de nuvem



On-premise

UX e recursos



UX consistente e recursos para leitores e escritores

Fácil de transformar os dados

- Operações suportadas
 - Concatenar múltiplos arquivos/tabelas
 - Filtro de coluna
 - Ordenar colunas
 - Renomear Colunas
 - Mapeamento do tipo de coluna
- Feedback instantâneo das mudanças efetuadas
- Acessível
 - em cada novo node leitor de tabela
 - via Table Manipulator

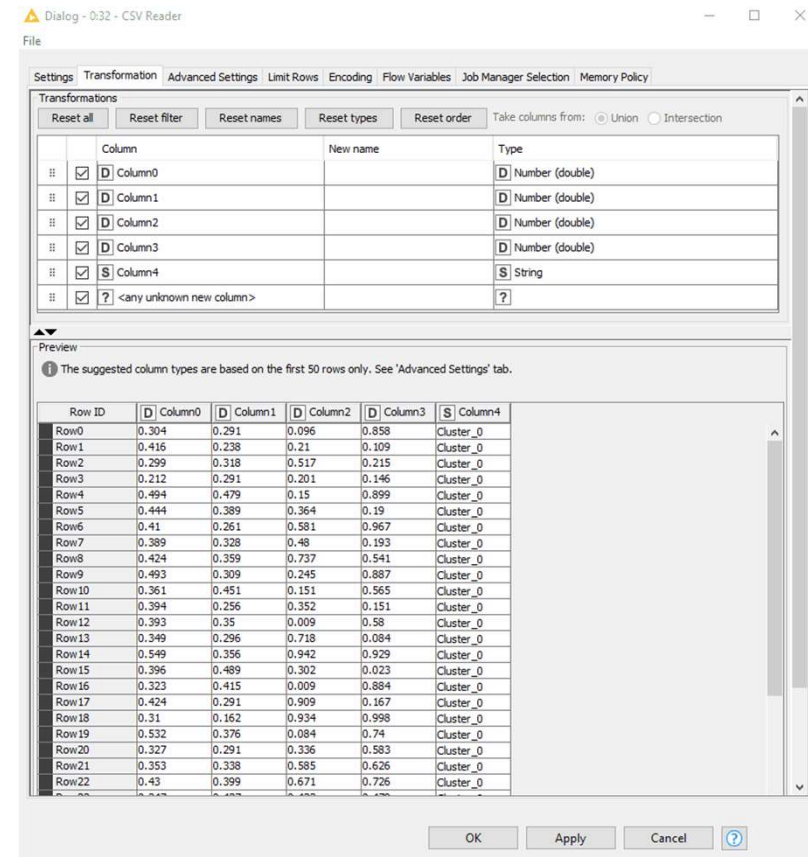


Table Manipulator

- Com base na nova estrutura do Table Reader
- Aplique transformações em qualquer número de tabelas de dados KNIME existentes

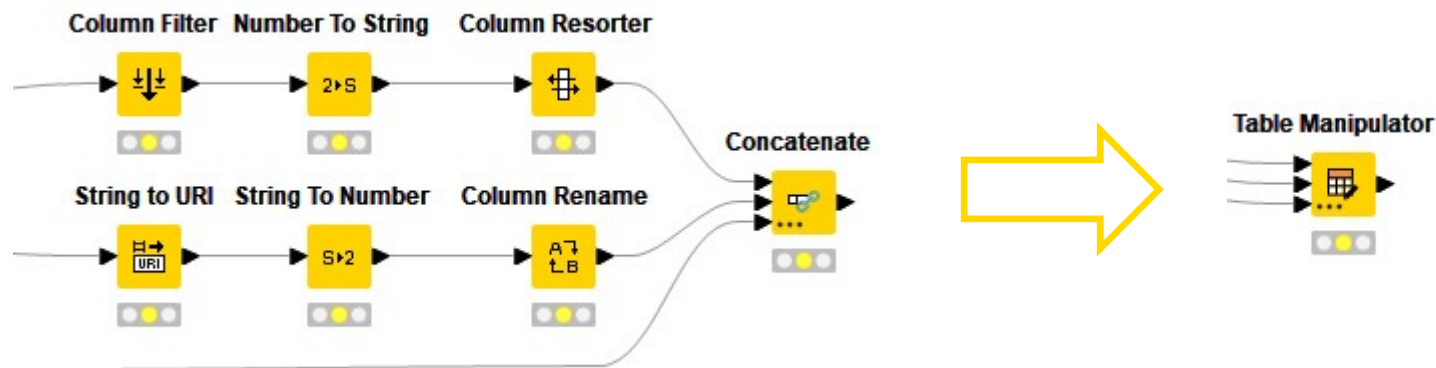


Table Manipulator

Dialog - 0:32 - CSV Reader

File

Settings Transformation Advanced Settings Limit Rows Encoding Flow Variables Job Manager Selection Memory Policy

Transformations

Reset all Reset filter Reset names Reset types Reset order Take columns from: Union Intersection

	Column	New name	Type
<input checked="" type="checkbox"/>	D Column0		D Number (double)
<input checked="" type="checkbox"/>	D Column1		D Number (double)
<input checked="" type="checkbox"/>	D Column2		D Number (double)
<input checked="" type="checkbox"/>	D Column3		D Number (double)
<input checked="" type="checkbox"/>	S Column4		S String
<input checked="" type="checkbox"/>	? <any unknown new column>		?

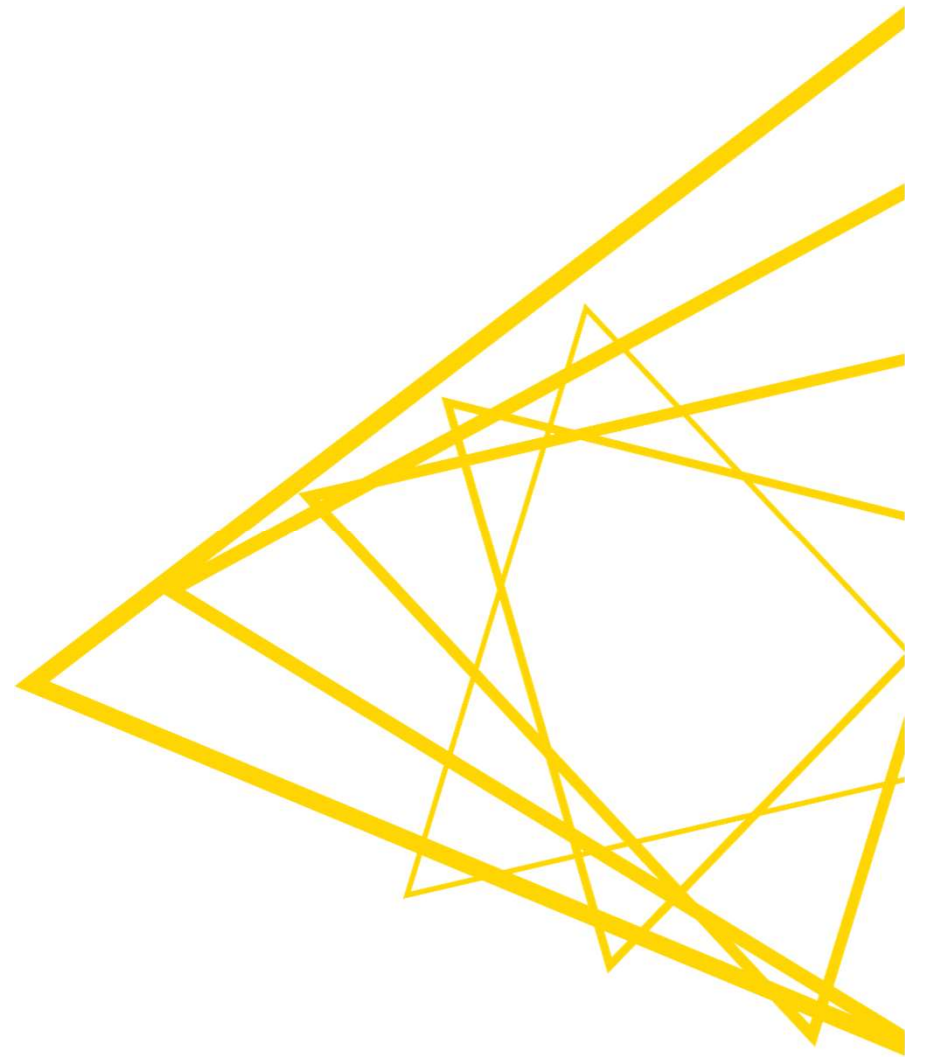
Preview

The suggested column types are based on the first 50 rows only. See 'Advanced Settings' tab.

Row ID	D Column0	D Column1	D Column2	D Column3	S Column4
Row0	0.304	0.291	0.096	0.858	Cluster_0
Row1	0.416	0.238	0.21	0.109	Cluster_0
Row2	0.299	0.318	0.517	0.215	Cluster_0
Row3	0.212	0.291	0.201	0.146	Cluster_0
Row4	0.494	0.479	0.15	0.899	Cluster_0
Row5	0.444	0.389	0.364	0.19	Cluster_0
Row6	0.41	0.261	0.581	0.967	Cluster_0
Row7	0.389	0.328	0.48	0.193	Cluster_0
Row8	0.424	0.359	0.737	0.541	Cluster_0
Row9	0.493	0.309	0.245	0.887	Cluster_0
Row10	0.361	0.451	0.151	0.565	Cluster_0
Row11	0.394	0.256	0.352	0.151	Cluster_0
Row12	0.393	0.35	0.009	0.58	Cluster_0
Row13	0.349	0.296	0.718	0.084	Cluster_0
Row14	0.549	0.356	0.942	0.929	Cluster_0
Row15	0.396	0.489	0.302	0.023	Cluster_0
Row16	0.323	0.415	0.009	0.884	Cluster_0
Row17	0.424	0.291	0.909	0.167	Cluster_0
Row18	0.31	0.162	0.934	0.998	Cluster_0
Row19	0.532	0.376	0.084	0.74	Cluster_0
Row20	0.327	0.291	0.336	0.583	Cluster_0
Row21	0.353	0.338	0.585	0.626	Cluster_0
Row22	0.43	0.399	0.671	0.726	Cluster_0

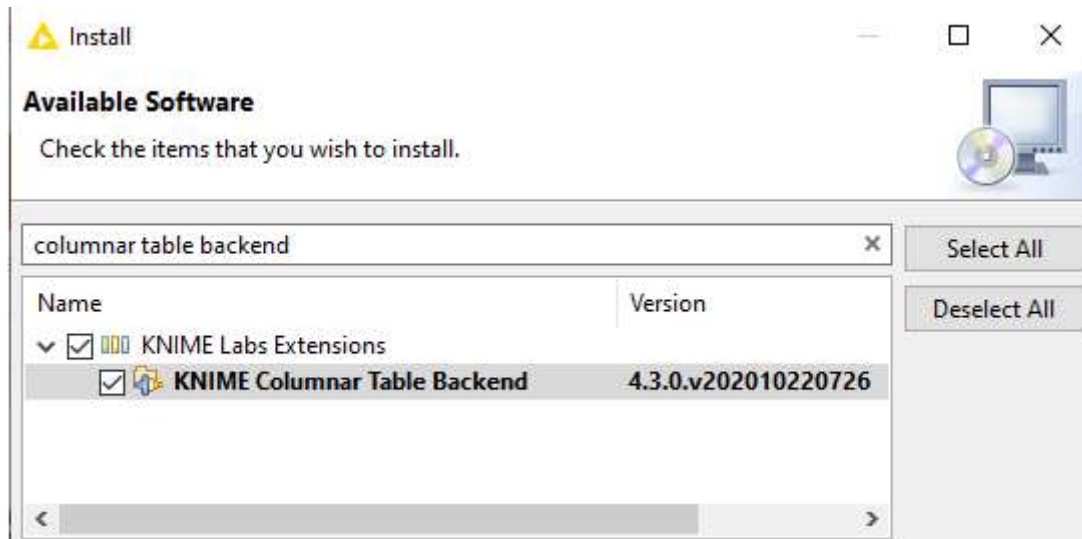
OK Apply Cancel ?

Desempenho



Columnar Table Backend for Fast Tables (KNIME Labs)

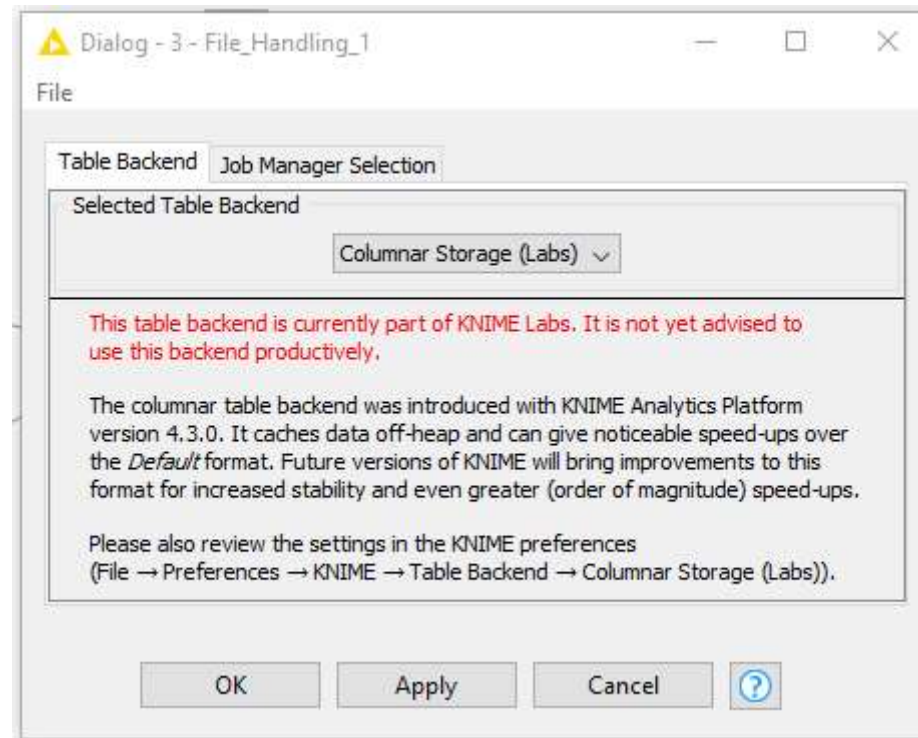
- Performance melhorada
- Eficiência de memória e no processamento in-memory



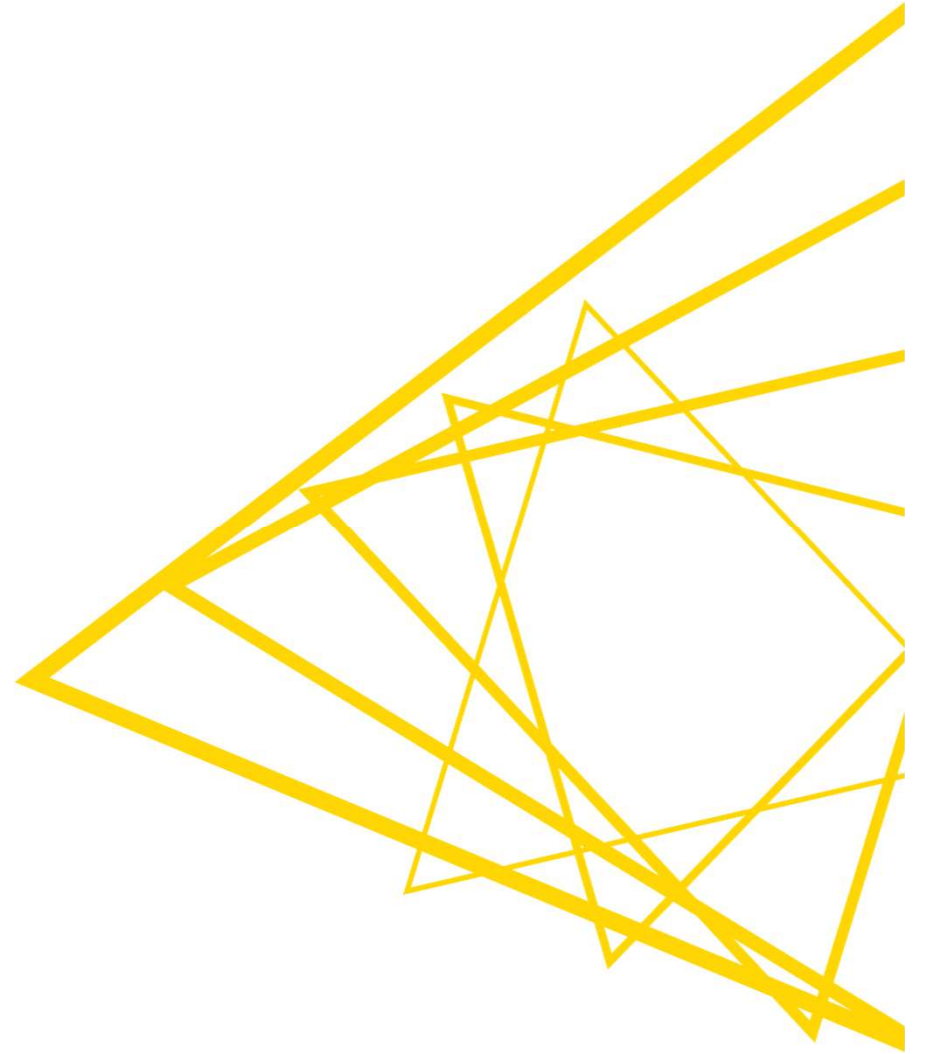
Blog: <https://www.knime.com/blog/improved-performance-with-new-table-backend>

Como usar já hoje?

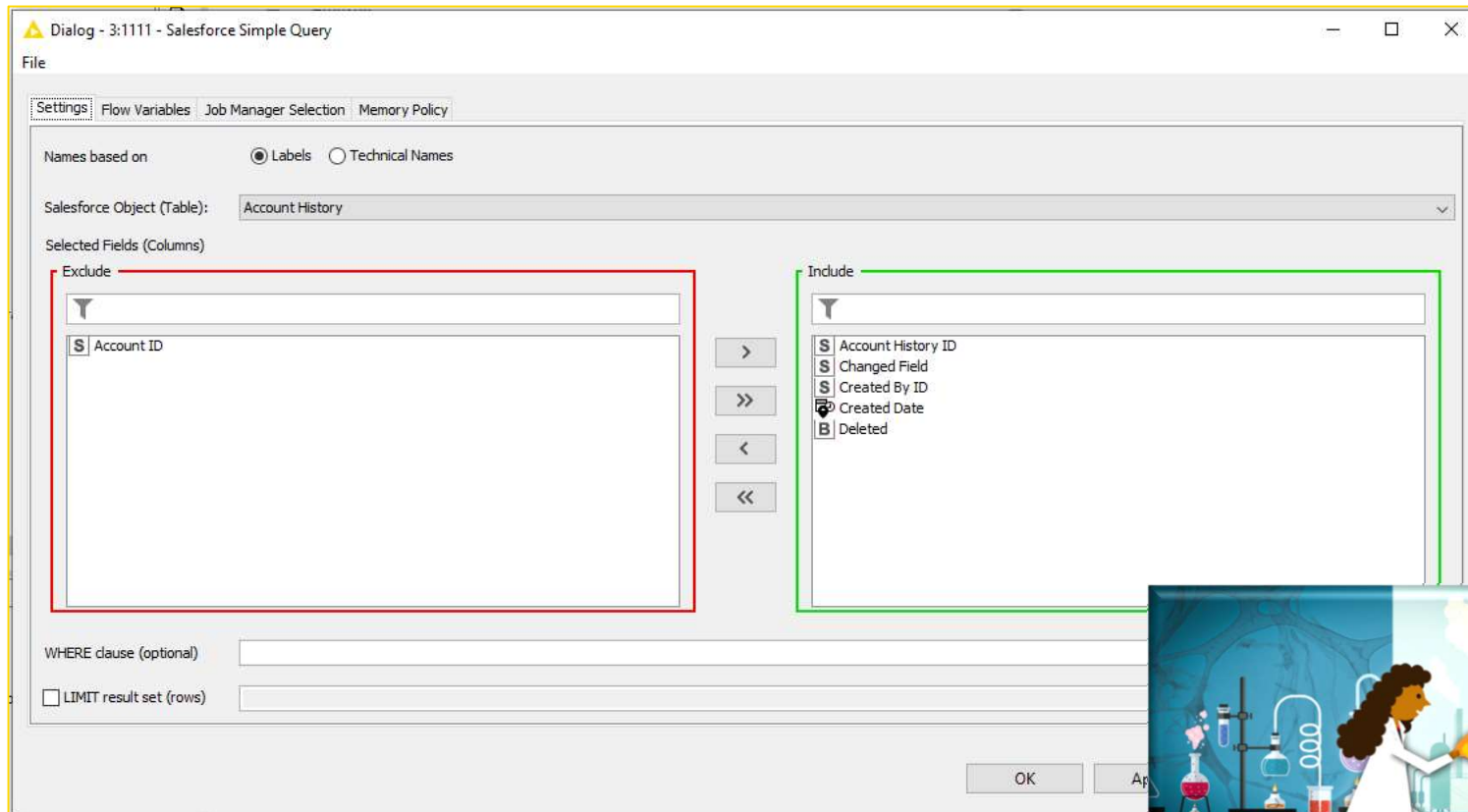
- Clique com o botão direito no Fluxo de Trabalho e selecione Configurar
- Selecione



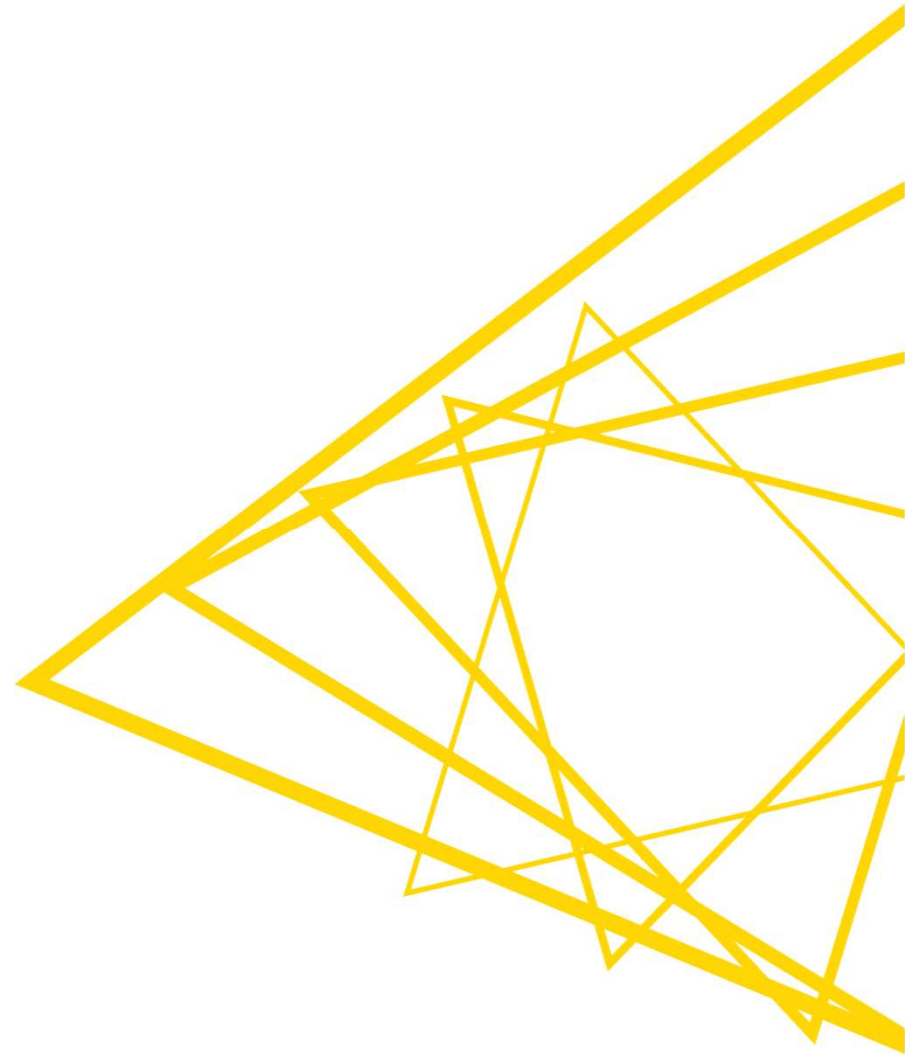
Extensões e Integrações



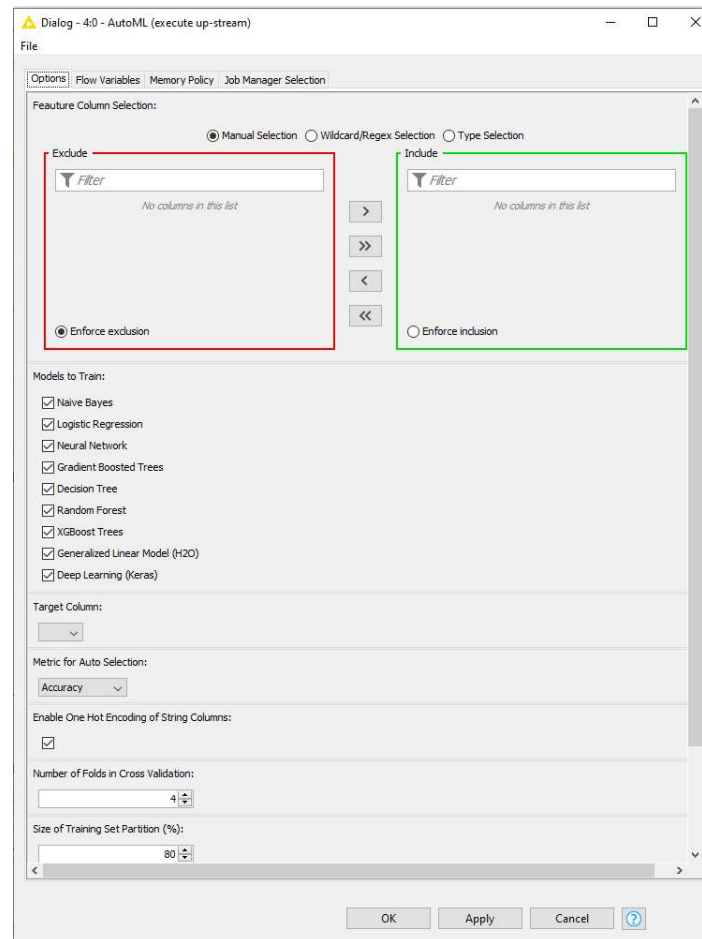
Novidades sobre Extensões e Integrações



Components



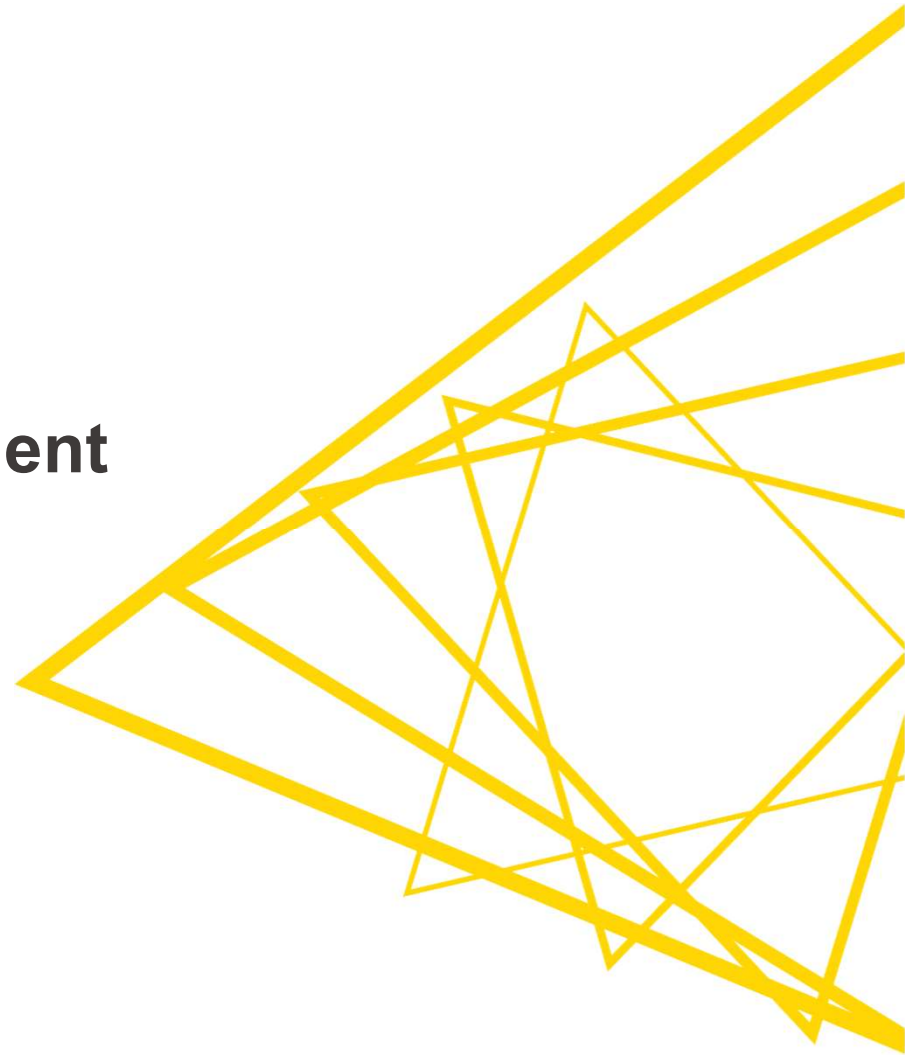
Components – Layout de Diálogo



Components – Layout de Diálogo

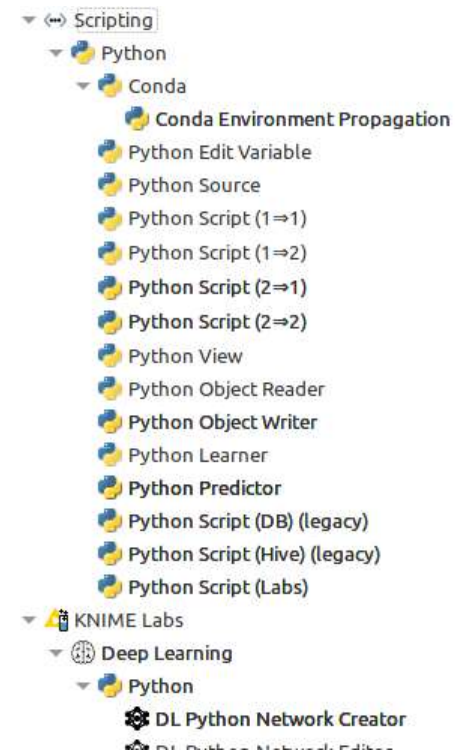
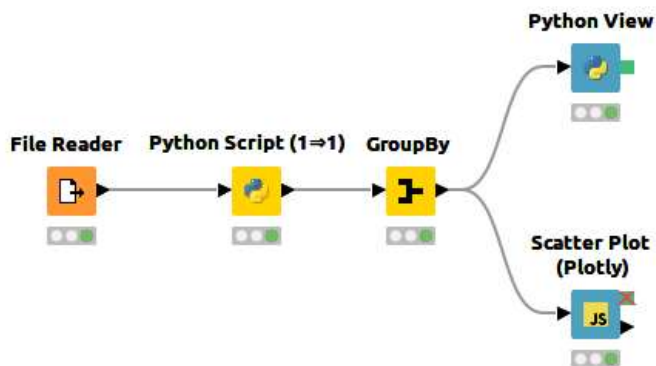


Python / Scripting / Deployment



Integração Python do KNIME

- Um conjunto de nodes para incluir scripts Python em um workflow do KNIME
- Scripting / Construção de modelo & Predição / Visualizar
- Troca de dados KNIME ↔ Python
- Apontar para um executor Python (Conda)



Conda Environment Propagation

- Defina um ambiente conda local e duplique este ambiente no server
 - Para executar em outro lugar, suas dependências de códigos Python devem estar presentes
 - Replicar o ambiente conda local de um fluxo de trabalho no servidor
- A estrutura prevê o ambiente com base na definição salva no fluxo de trabalho

Con

Include?	Name	Version	Build	Channel
<input checked="" type="checkbox"/>	mkl	2020.2	256	pkgs/main
<input checked="" type="checkbox"/>	mkl-service	2.3.0	py38he904b0f_0	pkgs/main
<input checked="" type="checkbox"/>	mkl_fft	1.2.0	py38h23d657b_0	pkgs/main
<input checked="" type="checkbox"/>	mkl_random	1.1.1	py38h0573a6f_0	pkgs/main
<input checked="" type="checkbox"/>	ncurses	6.2	he6710b0_1	pkgs/main
<input checked="" type="checkbox"/>	numpy	1.19.2	py38h54aff64_0	pkgs/main
<input checked="" type="checkbox"/>	numpy-base	1.19.2	py38hfa32c7d_0	pkgs/main
<input type="checkbox"/>	olefile	0.46	py_0	pkgs/main
<input checked="" type="checkbox"/>	openssl	1.1.1h	h7b6447c_0	pkgs/main
<input checked="" type="checkbox"/>	pandas	1.1.3	py38he6710b0_0	pkgs/main
<input checked="" type="checkbox"/>	paramiko	2.7.2	py_0	pkgs/main
<input type="checkbox"/>	pcre	8.44	he6710b0_0	pkgs/main
<input type="checkbox"/>	pillow	8.0.1	py38he98fc37_0	pkgs/main
<input checked="" type="checkbox"/>	pip	20.2.4	py38h06a4308_0	pkgs/main
<input checked="" type="checkbox"/>	protobuf	3.13.0.1	py38he6710b0_1	pkgs/main
<input checked="" type="checkbox"/>	pyparser	2.20	py_2	pkgs/main
<input checked="" type="checkbox"/>	pynacl	1.4.0	py38h7b6447c_1	pkgs/main
<input type="checkbox"/>	pyparsing	2.4.7	py_0	pkgs/main
<input type="checkbox"/>	pyqt	5.9.2	py38h05f1152_4	pkgs/main
<input checked="" type="checkbox"/>	python	3.8.5	h7579374_1	pkgs/main
<input checked="" type="checkbox"/>	python-dateutil	2.8.1	py_0	pkgs/main
<input checked="" type="checkbox"/>	pytz	2020.1	py_0	pkgs/main
<input type="checkbox"/>	qt	5.9.7	h5867ecd_1	pkgs/main
<input checked="" type="checkbox"/>	readline	8.0	h7b6447c_0	pkgs/main
<input checked="" type="checkbox"/>	setuptools	50.3.1	py38h06a4308_1	pkgs/main
<input checked="" type="checkbox"/>	sip	4.19.13	py38he6710b0_0	pkgs/main

Include all Exclude all Include only explicitly installed

Environment validation

Check name only

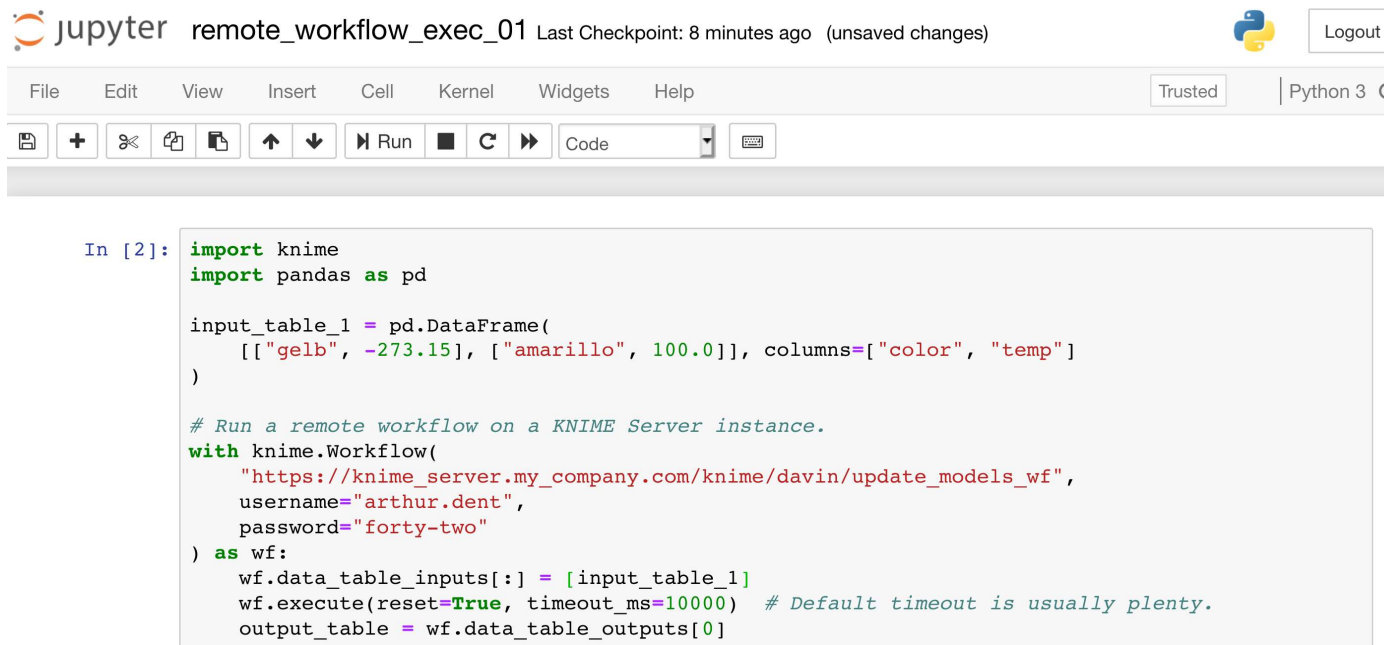
Check name and packages

Always overwrite existing environment

Outras mudanças dignas de nota: knimepy

- *knimepy* + KNIME Server

- O pacote *knimepy* Python adiciona capacidade para executar fluxos de trabalho remotos em um servidor
- Os dados locais podem ser usados como entrada; receba a saída de volta como pandas DataFrames
- Para usá-lo, forneça uma URL para o fluxo de trabalho no servidor



The screenshot shows a Jupyter Notebook interface with the following elements:

- Header: "jupyter remote_workflow_exec_01 Last Checkpoint: 8 minutes ago (unsaved changes)" with a Python logo and a "Logout" button.
- Menu: "File Edit View Insert Cell Kernel Widgets Help" with a "Trusted" status and "Python 3" environment.
- Toolbar: Includes icons for save, add, undo, redo, up/down arrows, run, and code editor.
- Code Cell: Contains the following Python code:

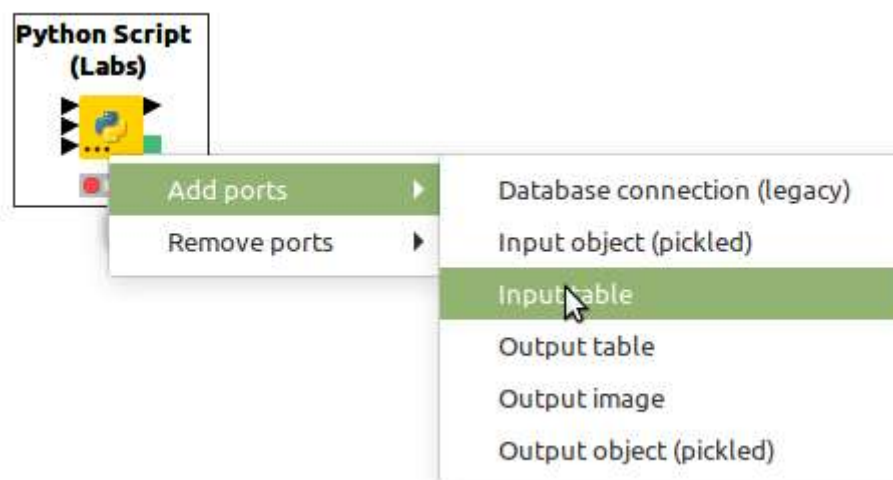
```
In [2]: import knime
import pandas as pd

input_table_1 = pd.DataFrame(
    [{"gelb", -273.15}, {"amarillo", 100.0}], columns=["color", "temp"]
)

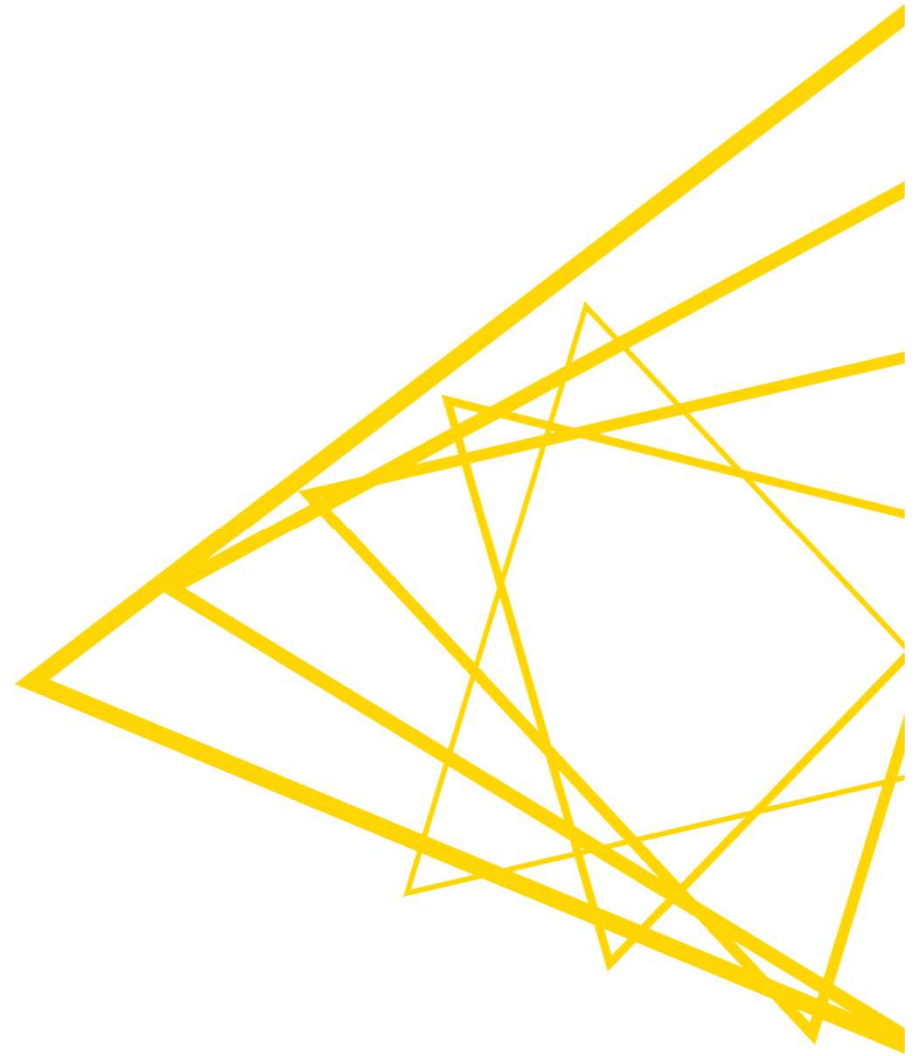
# Run a remote workflow on a KNIME Server instance.
with knime.Workflow(
    "https://knime_server.my_company.com/knime/davin/update_models_wf",
    username="arthur.dent",
    password="forty-two"
) as wf:
    wf.data_table_inputs[:] = [input_table_1]
    wf.execute(reset=True, timeout_ms=10000) # Default timeout is usually plenty.
    output_table = wf.data_table_outputs[0]
```

Outras mudanças dignas de nota : Portas Dinâmicas

- Dynamic Ports
 - Nodes de script Python suportam um número dinâmico de portas de entrada e saída
 - Substitui as limitações 1=>1, 1=>2, 2=>1, 2=>2
 - Adicione quantas portas de entrada ou saída desejar em um único node de script Python
 - Em seu script Python, cada porta de entrada fornece seu próprio Pandas DataFrame distinto

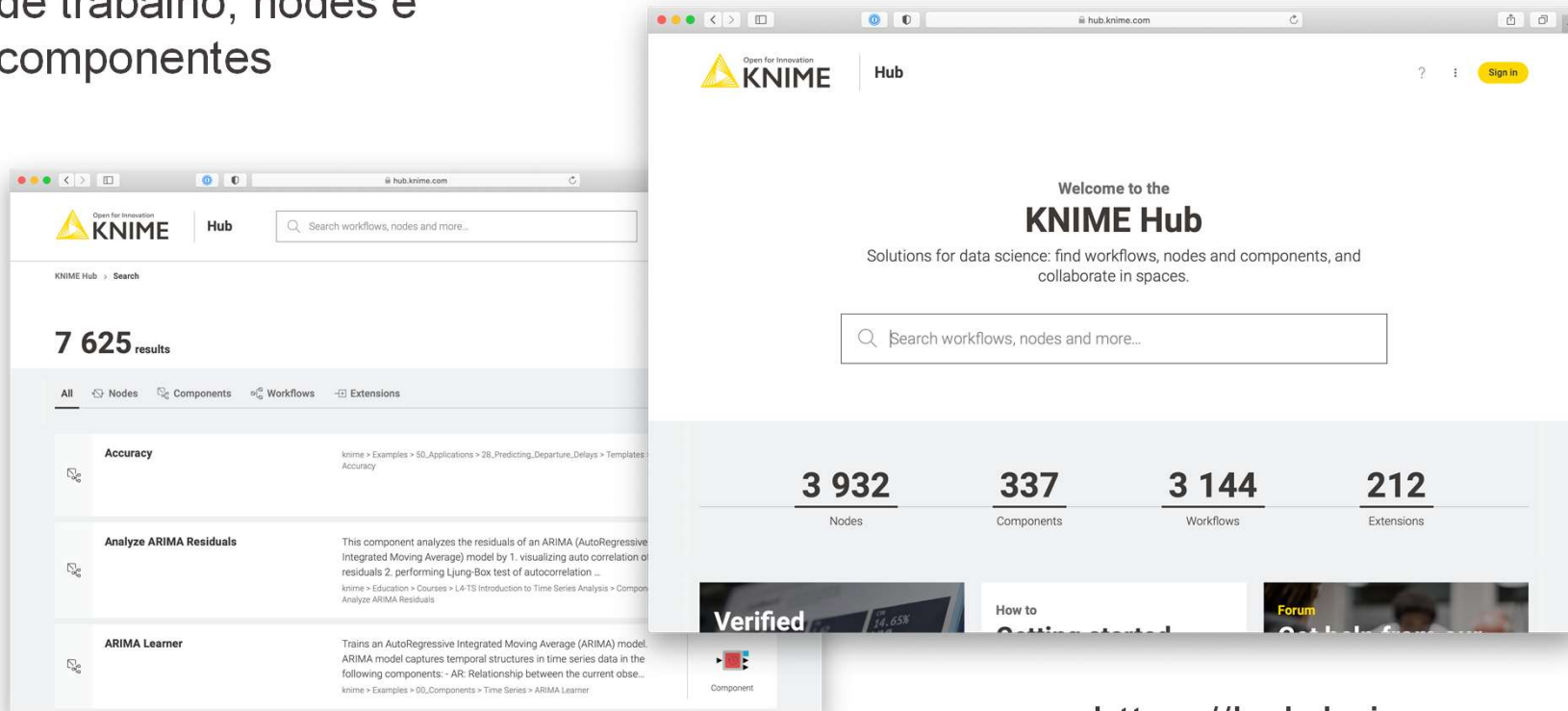


KNIME Hub



KNIME Hub – Soluções para Ciência de Dados

- Compartilhe e encontre fluxos de trabalho, nodes e componentes



<https://hub.knime.com>

KNIME Hub – Atualizações de Interface

The screenshot shows the user profile page for 'tschmidt' on the KNIME Hub. The page layout includes a top navigation bar with the KNIME logo, a search bar, and a user profile picture. The main content area features a user profile card with a profile picture, name, and role. Below this is a 'Your profile' section with various badges and a bio. A summary row displays statistics for workflows published, components published, extensions published, and likes received. The 'Your liked items' section lists two workflows: 'TweetKollidR' and 'COVID-19_Live_Visualization', each with a brief description and tags.

Open for Innovation
KNIME

Hub

Search workflows, nodes and more...

KNIME Hub > tschmidt

tschmidt
KNIME Team Member
Konstanz

Your profile

Leader Silver Anniversary Nice Topic

Real name Tobi Schmidt
Joined 2 years ago

Working on the KNIME Hub since 2018.

5 Workflows published | 1 Components published | 0 Extensions published | 2 Likes received

♥ Your liked items

TweetKollidR
Twitter visualisation network +2
A workflow to collect and visualise data from Twitter. Using the Twitter API, this workflow retrieves tweets that match specified search queries. It collates the results of repeated searches and prov...
angusveitch > Public > TweetKollidR

COVID-19_Live_Visualization
COVID-19 REST API +8
This workflow is connecting an API to get new data on the spreading of the virus COVID-19. Confirmed cases and deaths time series are visualized by country in different ways. The dashboard can be dis...
paolotamag > Public > COVID-19_Live_Visualization

KNIME Hub – Atualizações de Interface

The screenshot displays the KNIME Hub interface for a user named 'tschmidt'. The top navigation bar includes the KNIME logo, the word 'Hub', a search bar with the placeholder text 'Search workflows, nodes and more...', and a user profile picture. Below the navigation bar, the breadcrumb path 'KNIME Hub > tschmidt > Spaces' is visible. On the left side, there is a user profile card for 'tschmidt', identified as a 'KNIME Team Member' from 'Konstanz'. Below the profile card are three menu items: 'Overview', 'Extensions', and 'Spaces', with 'Spaces' being the active selection. The main content area is titled 'Your spaces' and features two project cards. The first card, 'Churn analysis', includes a description, a last update date of '4 Nov 2020', and statistics of 5 views, 1 share, and 0 downloads. The second card, 'Twitter trend project', includes a description, a last update date of '11 Nov 2020', and statistics of 0 views, 0 shares, and 0 downloads. At the bottom of the main content area, there is a 'Create new space' section with two options: 'Private space' and 'Public space'.

KNIME Hub - Atualizações de Interface

The screenshot displays the KNIME Hub interface. At the top, there is a navigation bar with the KNIME logo, the word 'Hub', and a search bar containing the text 'Search workflows, nodes and more...'. To the right of the search bar are icons for help and user profile. Below the navigation bar, the breadcrumb path reads 'KNIME Hub > tschmidt > Spaces > Churn analysis'. The main content area shows a 'Public space' with a title 'Churn analy' in a text box, accompanied by a yellow checkmark and a close button. Below the title, it indicates 'Last update: 4 Nov 2020' and shows '0' likes. A list of workflow components is displayed under the heading 'Home':

- Home
- data
- Clustering of customer groups
- Data connection
- Prepare data
- WebPortal dashboard

At the bottom of the workspace, there is a text box with a yellow checkmark and a close button, containing the following text:

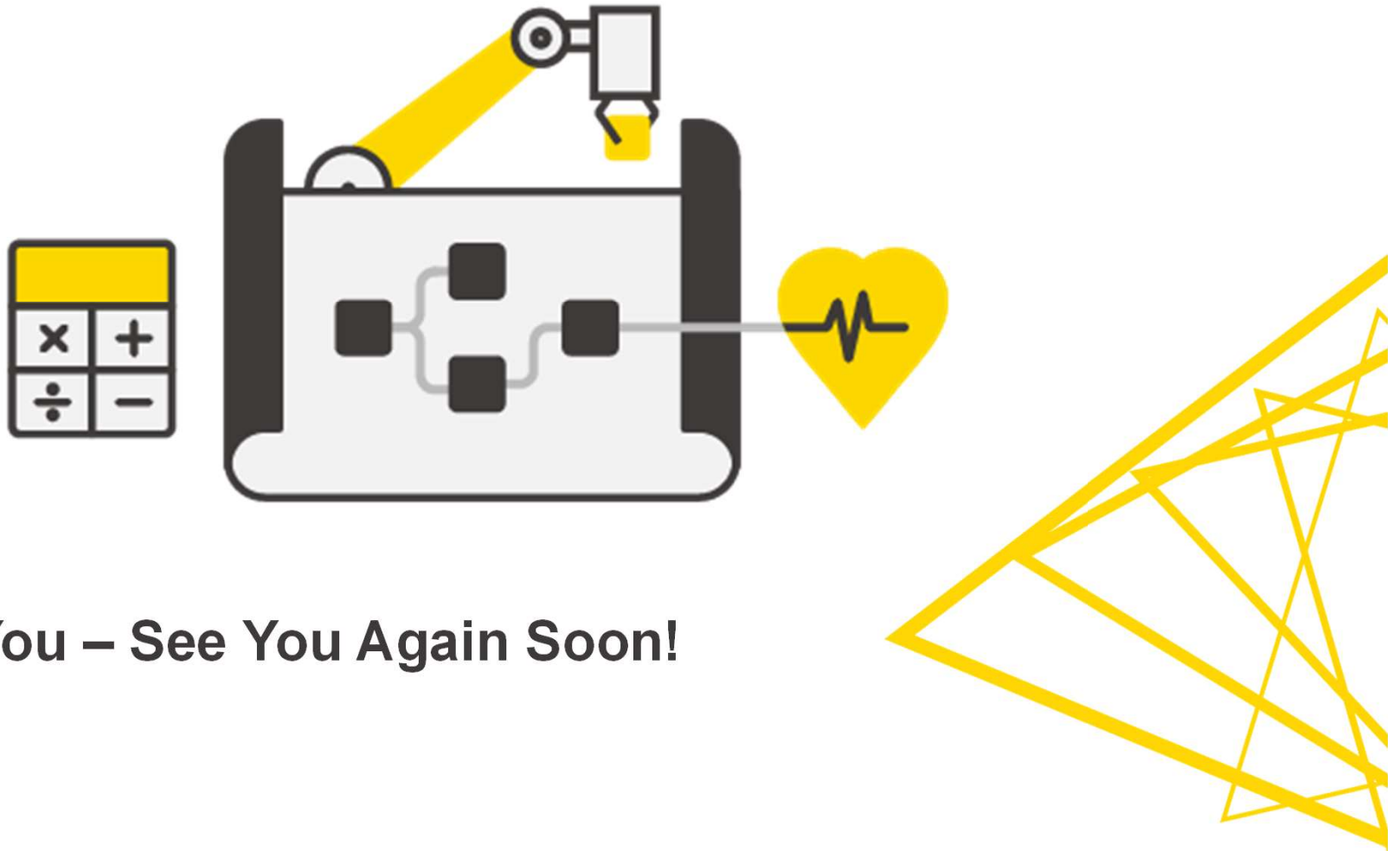
This project is the first step in our company's churn analysis. It includes data preparation and a first basic analysis. Please especial check

- the workflow "Prepare data"
- make sure to use the data connection from the components

All this should make it possible to present our results in a nice dashboard in the end.
Please contact me for

Integração direta no KNIME Analytics Platform

The screenshot displays the KNIME Analytics Platform interface. The central workspace shows a workflow titled "0: 01_ChurnAnalysis" with the following nodes: Excel Reader (XLS) (churn data: 3333 rows), Column Filter (deprecated) (Filter first 7 cols), Integer Configuration (Minimum k), Integer Configuration (Maximum k), Interval Loop Start (For each k), k-Means, Cluster Statistics, and Loop. A yellow arrow points from the search results panel to the workflow, with the text "Drag and drop" overlaid. The search results panel on the right shows a search for "churn" with 5 results, including "Reading and Pre-Processing Data_v1" and "Customer Count and Churn Rate by Time-based Cohort". The bottom right panel shows the Node Monitor for the "Column Filter (deprecated)" node, indicating it is in the "EXECUTED" state.



Thank You – See You Again Soon!