

# Introdução ao KNIME

## Palestras do DCC

Geraldo Xexéo<sup>1,2</sup>

<sup>1</sup>Departamento de Ciências da Computação

<sup>2</sup>Programa de Engenharia de Sistemas e Computação

Palestra Remota para o DCC, Junho 2020

# Agenda

- 1 O que é o KNIME
- 2 Workflows
- 3 Exemplos
- 4 BIRT – Basic Intelligence Report Tool
- 5 Conclusão

# Onde Estamos?

## 1 O que é o KNIME

## 2 Workflows

## 3 Exemplos

## 4 BIRT – Basic Intelligence Report Tool

## 5 Conclusão

# Apresentação do KNIME

- Pronúncia “NÁIMI”
- 2 Ferramentas de *Data Science*
  - KNIME Analytics Platform
    - Open Source
    - Desenvolvimentos de sistemas de *Data Science*
  - KNIME Server
    - Comercial
    - Colaboração, gerência, automação e *deployment* de *workflows* de *Data Science* como aplicações e serviços

# KNIME Analytics Platform

- Baseada no Eclipse
- Implementa vários algoritmos
- Programável e extensível
  - Java, Python, R
- Gera relatórios e gráficos
- Possui plug-ins de terceiros

O que é o KNIME  
Workflows  
Exemplos  
BIRT – Basic Intelligence Report Tool  
Conclusão

Apresentação do KNIME  
KNIME Analytics Platform  
A interface do KNIME  
O que é um workflow

# A interface do KNIME

The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a workflow named "Palestra-Exemplo-Workflow-1" with three nodes: "Excel Reader (XLS)", "Row Filter", and "CSV Writer". The workflow is connected by arrows, indicating a sequential process. Below the nodes, there are labels: "Ler arquivo MS-Covid" under Excel Reader, "Papa apenas dados do Brasil" under Row Filter, and "Salva em arquivo CSV" under CSV Writer.

The left sidebar contains the "KNIME Explorer" pane, showing a tree view of the project structure. It includes folders for "My KNIME-Hub", "EXAMPLES", "LOCAL", and "Palestra Knime". The "Palestra Knime" folder is expanded, showing the "Palestra-Exemplo-Workflow-1" workflow.

The right sidebar contains the "Description" pane, which provides metadata for the selected workflow. It includes fields for "Title", "Description", "Tags", "Links", "Creation Date", and "Author".

The bottom of the interface features the "Outline" pane, which shows a hierarchical view of the workflow nodes, and the "Console" pane, which displays the execution log and any warnings or errors.

The console output shows the following messages:

```
KNIME Console
*** welcome to KNIME Analytics Platform v4.1.3.v202005121100 ***
*** Copyright by KNIME AG, Zurich, Switzerland ***
*****
Log file is located at: F:\Users\Gerald\knime-workspace\.metadata\knime.log
WARN Excel Reader (XLS) 0:4 The file/directory "F:\Github\Palestra KNIME\Data\HIST_PAINEI_COVID08_27mai2020.xls"
WARN Excel Reader (XLS) 0:4 failed to apply settings: Loading input file...
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Output file "F:\Github\Palestra-KNIME\Data\Resultado-1.csv" exists and will be overw
```

O que é o KNIME  
Workflows  
Exemplos  
BIRT – Basic Intelligence Report Tool  
Conclusão

Apresentação do KNIME  
KNIME Analytics Platform  
A interface do KNIME  
O que é um workflow

# A interface do KNIME

The screenshot displays the KNIME Analytics Platform interface. The central workspace shows a workflow titled "Palestra-Exemplo-Workflow-1" with three nodes: "Excel Reader (XLS)", "Row Filter", and "CSV Writer". The workflow is connected by arrows, indicating a sequential process. Below the nodes, there are descriptive labels: "Ler arquivo MS-Covid" for the Excel Reader, "Filtro apenas dados do Brasil" for the Row Filter, and "Salva em arquivo CSV" for the CSV Writer.

A blue box with the text "Área do Workflow" is overlaid on the central workspace. The interface includes a menu bar (File, Edit, View, Node, Help), a toolbar, and a sidebar with various panels: "KNIME Explorer" (showing project structure), "Favorite Nodes" (personal and frequently used), "Node Repository" (categorized nodes), "Outline" (hierarchical view of the workflow), and "Console" (log output).

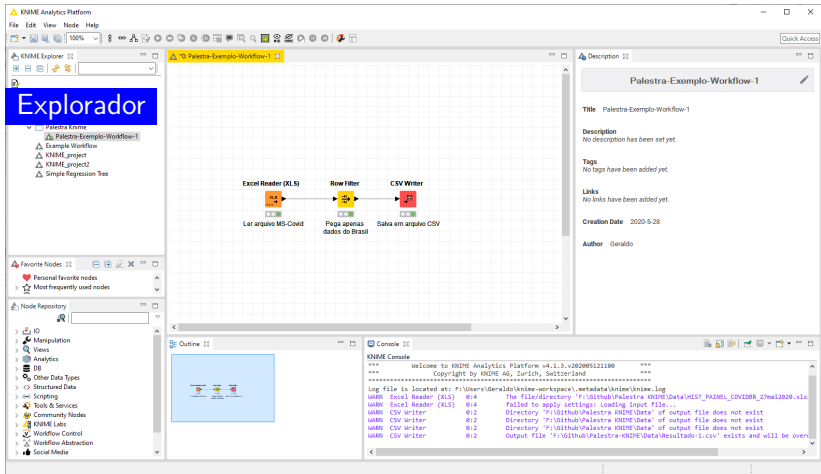
The "Description" panel on the right provides details about the workflow, including its title, description, tags, links, creation date, and author.

The "Console" panel at the bottom right shows the log output, including the KNIME version (v4.1.3.v202005121100) and the path to the log file.

O que é o KNIME  
Workflows  
Exemplos  
BIRT – Basic Intelligence Report Tool  
Conclusão

Apresentação do KNIME  
KNIME Analytics Platform  
A interface do KNIME  
O que é um workflow

# A interface do KNIME





O que é o KNIME  
Workflows  
Exemplos  
BIRT – Basic Intelligence Report Tool  
Conclusão

Apresentação do KNIME  
KNIME Analytics Platform  
A interface do KNIME  
O que é um workflow

# A interface do KNIME

The screenshot displays the KNIME Analytics Platform interface. On the left, the 'KNIME Explorer' pane shows a project tree with 'Palestra-Exemplo-Workflow-1' selected. Below it, the 'Favorite Nodes' pane lists 'Personal favorite nodes' and 'Most frequently used nodes'. The main workspace shows a workflow with three nodes: 'Excel Reader (XLS)', 'Row Filter', and 'CSV Writer'. The 'Excel Reader (XLS)' node is connected to the 'Row Filter' node, which is connected to the 'CSV Writer' node. The 'Excel Reader (XLS)' node has a description 'Lar arquivo MS-Covid'. The 'Row Filter' node has a description 'Papa apenas dados do Brasil'. The 'CSV Writer' node has a description 'Salva em arquivo CSV'. On the right, the 'Description' pane shows the workflow's title 'Palestra-Exemplo-Workflow-1', description, tags, links, creation date (2020-5-28), and author (Geraldo). At the bottom, the 'Console' pane shows the KNIME Console output, including a welcome message and log file information.

**Nós disponíveis**

- Analytics
- DB
- Other Data Types
- Structured Data
- Tools & Services
- Community Nodes
- KNIME Labs
- Workflow Control
- Workflow Abstraction
- Social Media

```
KNIME Console
*** welcome to KNIME Analytics Platform v4.1.3.v202005121100 ***
*** Copyright by KNIME AG, Zurich, Switzerland ***
*****
Log file is located at: F:\Users\Geraldo\knime-workspace\.metadata\knime.log
WARN Excel Reader (XLS) 0:4 The file/directory "F:\Github\Palestra KNIME\data\HIST_PAINEI_COVID08_27mai2020.xls"
WARN Excel Reader (XLS) 0:4 failed to apply settings: Loading input file...
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\data" of output file does not exist
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\data" of output file does not exist
WARN CSV Writer 0:2 Output file "F:\Github\Palestra-KNIME\data\Resultado-1.csv" exists and will be over
```

O que é o KNIME  
Workflows  
Exemplos  
BIRT – Basic Intelligence Report Tool  
Conclusão

Apresentação do KNIME  
KNIME Analytics Platform  
A interface do KNIME  
O que é um workflow

# A interface do KNIME

The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a workflow titled "Palestra-Exemplo-Workflow-1" with three nodes: "Excel Reader (XLS)", "Row Filter", and "CSV Writer". The workflow is connected by arrows, indicating a sequential process. Below the nodes, there are labels: "Lar arquivo MS-Covid" under the Excel Reader, "Papa apenas dados do Brasil" under the Row Filter, and "Salva em arquivo CSV" under the CSV Writer.

On the right side, the "Description" panel is visible, showing the title "Palestra-Exemplo-Workflow-1" and a large blue box with the word "Descrição". Below this, there are sections for "Title", "Description", "Tags", "Links", "Creation Date", and "Author".

The bottom of the interface features a "Console" panel with the following text:

```
KNIME Console
*** welcome to KNIME Analytics Platform v4.1.3.v202005121100 ***
*** Copyright by KNIME AG, Zurich, Switzerland ***
*****
Log file is located at: F:\Users\Gerald\knime-workspace\.metadata\knime.log
WARN Excel Reader (XLS) 0:4 The file/directory "F:\Github\Palestra KNIME\Data\HIST_PAINEI_COVID08_27mai2020.xls"
WARN Excel Reader (XLS) 0:4 failed to apply settings: Loading input file...
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Output file "F:\Github\Palestra-KNIME\Data\Resultado-1.csv" exists and will be overw
```

O que é o KNIME  
Workflows  
Exemplos  
BIRT – Basic Intelligence Report Tool  
Conclusão

Apresentação do KNIME  
KNIME Analytics Platform  
A interface do KNIME  
O que é um workflow

# A interface do KNIME

The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a workflow titled "Palestra-Exemplo-Workflow-1" with three nodes: "Excel Reader (XLS)", "Row Filter", and "CSV Writer". The workflow is connected by arrows, indicating a sequential process. Below the nodes, there are labels: "Lar arquivo MS-Covid" under the Excel Reader, "Papa apenas dados do Brasil" under the Row Filter, and "Salva em arquivo CSV" under the CSV Writer.

The left sidebar contains the "KNIME Explorer" pane, showing a tree view of the project structure. Below it is the "Favorite Nodes" pane, and at the bottom is the "Node Repository" pane, which lists various nodes categorized by IO, Manipulation, Views, Analytics, DB, Other Data Types, Structured Data, Scripting, Tools & Services, Community Nodes, KNIME Labs, Workflow Control, Workflow Abstraction, and Social Media.

The right sidebar shows the "Description" pane for the selected workflow, displaying metadata such as Title, Description, Tags, Links, Creation Date, and Author.

The bottom of the interface features an "Outline" pane and a "Console" pane. The "Console" pane shows the KNIME log output, including the welcome message and the path to the log file. The log output indicates that the file directory "F:\Users\Gerald\KNIME\data\KNIME\HIST\_PAINEL\_COVID\_2020\_01\_27\2020\_01\_27" does not exist, and the output file "F:\Users\Gerald\KNIME\data\KNIME\HIST\_PAINEL\_COVID\_2020\_01\_27\2020\_01\_27\Resultado-1.csv" will be created.

O que é o KNIME  
Workflows  
Exemplos  
BIRT – Basic Intelligence Report Tool  
Conclusão

Apresentação do KNIME  
KNIME Analytics Platform  
A interface do KNIME  
O que é um workflow

# A interface do KNIME

The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a workflow titled "Palestra-Exemplo-Workflow-1" with three nodes: "Excel Reader (XLS)", "Row Filter", and "CSV Writer". The workflow is connected by arrows, indicating a sequential process. Below the workflow, the "Outline" pane shows a tree view of the workflow's structure. The "Console" pane at the bottom right displays the execution log, which includes the following text:

```
KNIME Console
*** welcome to KNIME ***
*** Copying KNIME ***
Log file is located at: F:\Users\Gerald\knime-workspace\workspace\knime\knime.log
WARN Excel Reader (XLS) 0:4 The file/directory "F:\Github\Palestra KNIME\Data\HIST_PAINEI_COVID08_27mai2020.xls"
WARN Excel Reader (XLS) 0:4 Failed to apply settings: Loading input file...
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Output file "F:\Github\Palestra-KNIME\Data\Resultado-1.csv" exists and will be over
```

The interface also includes a "Description" pane on the right, which provides details about the workflow, such as its title, description, tags, links, creation date, and author.

# O que é um workflow

Um workflow, ou fluxo de trabalho, é uma forma de descrever uma cadeia de processamento por meio de um grafo onde nós representam processos e arestas representam o fluxo de dados entre esses processos.

## Excel Reader (XLS)



Ler arquivo MS-Covid

## Row Filter



Pega apenas dados do Brasil

## CSV Writer



Salva em arquivo CSV

# Onde Estamos?

1 O que é o KNIME

2 Workflows

3 Exemplos

4 BIRT – Basic Intelligence Report Tool

5 Conclusão

# Entendendo um Workflow

- Um workflow é composto de nós
- Cada nó faz uma atividade
- Nós são divididos em tipos
- Nós tem portões de entrada e saída, de tipos diferentes
- Portões são conectados por fluxos, passando os dados e informações sobre eles

## Excel Reader (XLS)



Ler arquivo MS-Covid

## Row Filter



Pega apenas  
dados do Brasil

## CSV Writer



Salva em arquivo CSV

## Um nó básico

- Tipo do nó
- Desenho indicativo do nó
- Porta de entrada
- Porta de saída
- Status do nó
- Nome do nó

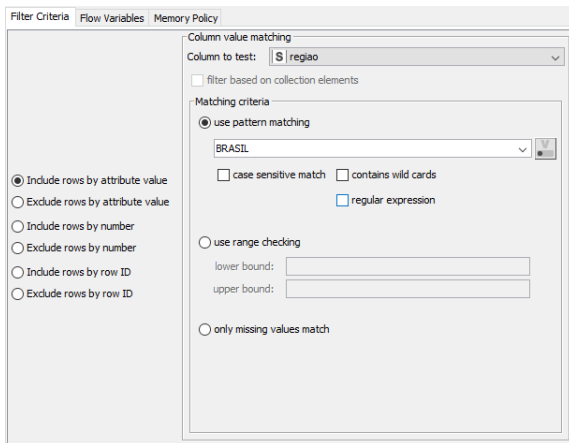




## Status do Nó

- Não configurado
- Pronto para executar
- Executado

# Configuração de um Nó



# Tipos de Nós

- 1 Leitura
- 2 Escrita / *Deploy*
- 3 Transformação
- 4 Análise, Aprendizado
- 5 Exploração
- 6 Predição
- 7 Controle

**File Reader**



Node 1

**CSV Writer**



Node 2

**Auto-Binner**



Node 3

**Decision Tree Learner**



Node 4

**Histogram**



Node 5

**Decision Tree Predictor**



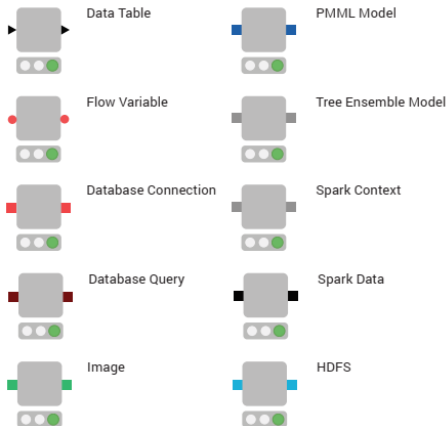
Node 6

**Counting Loop Start**



Node 7

## Tipos de Porta (mais comuns)



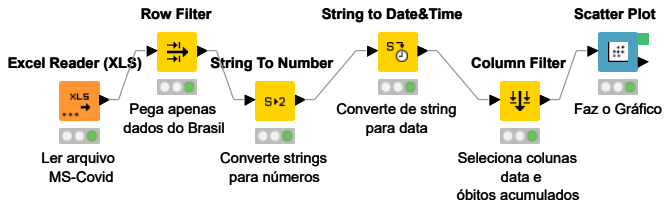
# Construindo um Workflow

- Coloque nós no mapa
- Configure os nós
- Execute
- Repita até alcançar o desejado

# Onde Estamos?

- 1 O que é o KNIME
- 2 Workflows
- 3 Exemplos**
- 4 BIRT – Basic Intelligence Report Tool
- 5 Conclusão

# Visualização de Curvas



# Metanodes

## Excel Reader (XLS)



## Seleciona Colunas

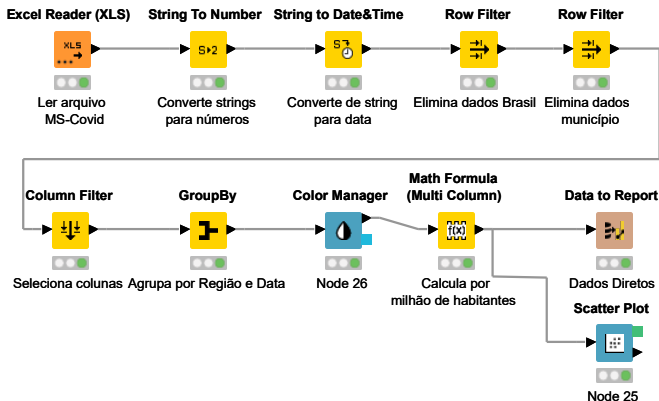


## Scatter Plot

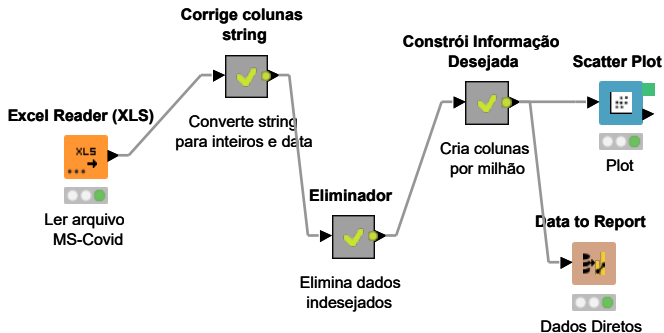




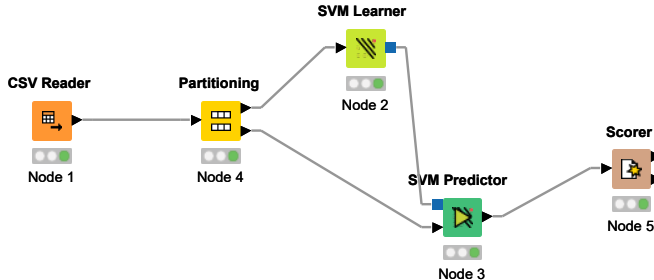
# Estudando o Covid-19



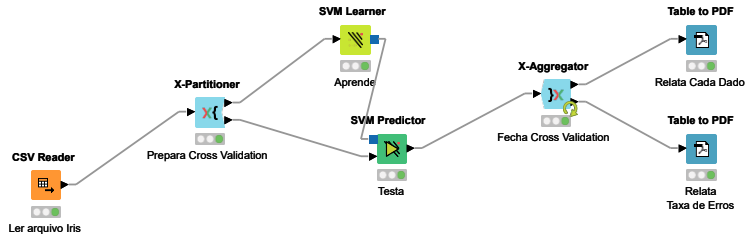
# Simplificando o Modelo



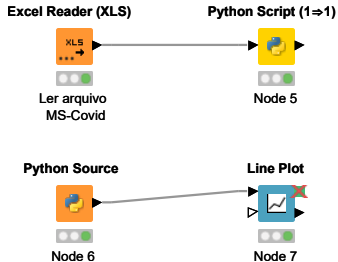
# Aprendizado de Máquina



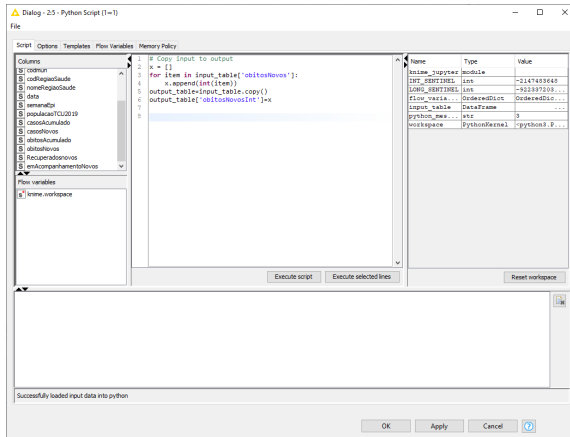
# Cross Validation



# Python Script



## Janela do Python

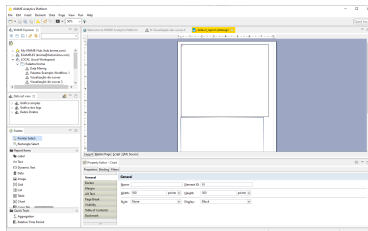


# Onde Estamos?

- 1 O que é o KNIME
- 2 Workflows
- 3 Exemplos
- 4 BIRT – Basic Intelligence Report Tool
- 5 Conclusão

# O que é BIRT

Software open source usado para criar relatórios, baseado no resultado dos workflows. Podem ser gerados em PDF, HTML e outros formatos.





# Como usar BIRT

- instalar (instalar plug-in no )

- dentro de um workflow, clicar



## Como usar BIRT II

- defina o layout da página
- defina a master page
- usos mais avançados com scripts...

# Onde Estamos?

- 1 O que é o KNIME
- 2 Workflows
- 3 Exemplos
- 4 BIRT – Basic Intelligence Report Tool
- 5 Conclusão**

## Conclusão

KNIME é um ambiente de desenvolvimento (e execução) de processos de Data Science de alta flexibilidade, podendo ser extendido por meio de plug-ins, scripts ou programas externos. Aviso: a documentação é de qualidade média. Mas existem muitos vídeos de treinamento.

É uma ótima ferramenta para Projetos Finais de Data Science.

# Obrigado!

## Contato



UNIVERSIDADE FEDERAL  
DO RIO DE JANEIRO



INSTITUTO DE MATEMÁTICA  
Universidade Federal do Rio de Janeiro



DEPARTAMENTO DE  
CIÊNCIA DA  
COMPUTAÇÃO  
- UFRJ -



LINE  
Laboratório de Tratamento da  
Informação Não Estruturada



COPPE  
Instituto Alberto Luis Coimbra de  
Pós-Graduação e Pesquisa de Engenharia  
UFRJ



PESCE  
Programa de Engenharia  
de Sistemas e Computação



LUDDES  
Ludologia, Engenharia e Simulação



Geraldo Xexéo

[xexeo@cos.ufrj.br](mailto:xexeo@cos.ufrj.br)

[gxexeo@gmail.com](mailto:gxexeo@gmail.com)

Este obra está licenciado com uma Licença [Creative Commons Atribuição-NãoComercial-SemDerivações 4.0 Internacional](#).