

# 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

# Onde Estamos?

1 O que é o KNIME

2 Workflows

3 Exemplos

4 BIRT – Basic Intelligence Report Tool

# Apresentação do KNIME

- Pronúncia “NAIME”
- 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

# 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 descriptive labels: "Ler arquivo MS-Covid" for the Excel Reader, "Pega apenas dados do Brasil" for the Row Filter, and "Salva em arquivo CSV" for the CSV Writer.

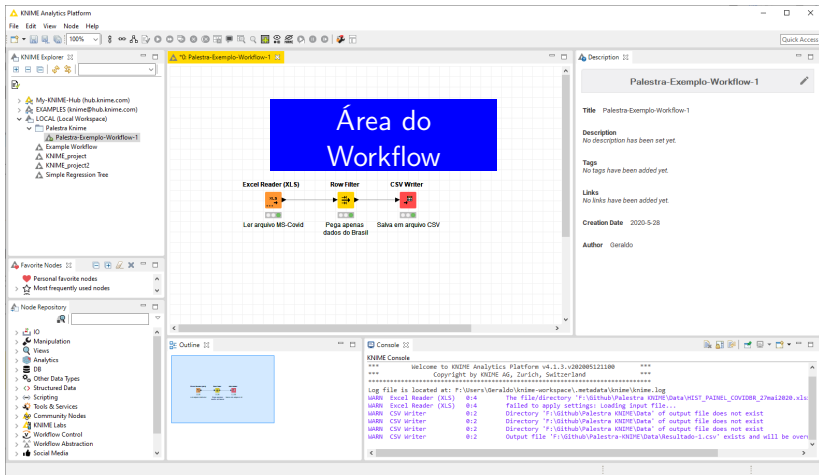
The left sidebar contains the "KNIME Explorer" and "Favorite Nodes" panels. The "KNIME Explorer" shows a tree structure of projects and workflows, including "My-KNIME-Hub", "EXAMPLES", "LOCAL (Local Workspace)", and "Palestra KNIME". The "Favorite Nodes" panel shows a list of nodes categorized by "Personal favorite nodes" and "Most frequently used nodes".

The right sidebar shows the "Description" panel for the selected workflow, displaying its title, description, tags, links, creation date, and author.

The bottom panel shows the "Console" window, which displays the KNIME log output. The log includes the following information:

```
KNIME Console
*** Welcome to KNIME Analytics Platform v4.13.0(202005121100) ***
*** 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_COVID2019_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 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...
```

# A interface do KNIME



# A interface do KNIME

## Explorador

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, descriptive text reads: "Ler arquivo MS-Covid", "Pega apenas dados do Brasil", and "Salva em arquivo CSV".

The left sidebar contains the "KNIME Explorer" and "Favorite Nodes" panels. The "KNIME Explorer" shows a tree structure with "My KNIME-Hub (hub.knime.com)" and "EXAMPLES (knime@hub.knime.com)". The "Favorite Nodes" panel lists "Personal favorite nodes" and "Most frequently used nodes". The "Node Repository" panel shows a search bar and a list of 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" panel for the selected workflow, including fields for "Title", "Description", "Tags", "Links", "Creation Date", and "Author".

The bottom panel shows the "Console" output, which includes a welcome message and a log of the workflow execution. The log indicates that the "Excel Reader (XLS)" node failed to apply settings, and the "CSV Writer" node failed to create the output file "F:\Github\Palestra KNIME\Data\Resultado-1.csv".



# 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 as follows: "Excel Reader (XLS)" → "Row Filter" → "CSV Writer". Below each node, there is a description: "Ler arquivo MS-Covid", "Pega apenas dados do Brasil", and "Salva em arquivo CSV" respectively.

The left sidebar contains the "KNIME Explorer" and "Favorite Nodes" panels. The "KNIME Explorer" shows the project structure, including "My KNIME-Hub", "EXAMPLES", and "LOCAL (Local Workspace)". The "Favorite Nodes" panel shows "Personal favorite nodes" and "Most frequently used nodes".

The right sidebar contains the "Description" panel, which shows the title "Palestra-Exemplo-Workflow-1", a description "No description has been set yet.", tags "No tags have been added yet.", links "No links have been added yet.", creation date "2020-5-28", and author "Geraldo".

The bottom panel shows the "Console" output, which includes the following text:

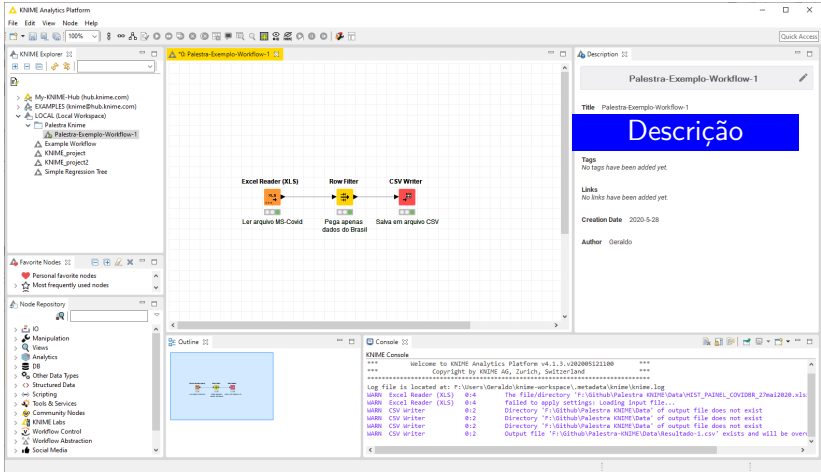
```

KNIME Console
=====
Welcome to KNIME Analytics Platform v4.1.3.v202005121100
Copyright by KNIME AG, Zurich, Switzerland
=====
Log file is located at: F:\Users\Geraldo\knime-workspace\metadata\knime\log
WARN: Excel Reader (XLS): 0:4 The file/directory 'F:\Github\Palestra KNIME\Data\HIST_PAINEI_COVID2019_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 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
  
```

Nós disponíveis

BIRT – Basic Intelligence Report Tool

# A interface do KNIME



# 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 descriptive labels: "Ler arquivo MS-Covid" for the Excel Reader, "Pega apenas dados do Brasil" for the Row Filter, and "Salva em arquivo CSV" for the CSV Writer.

On the left side, the "KNIME Explorer" pane shows the project structure, including "My-KNIME-Hub", "EXAMPLES", and "LOCAL (Local Workspace)". The "Favorite Nodes" pane shows "Personal favorite nodes" and "Most frequently used nodes". The "Node Repository" pane shows various node categories like "IO", "Manipulation", "Views", "Analytics", "DB", "Other Data Types", "Scripting", "Tools & Services", "Community Nodes", "KNIME Labs", "Workflow Control", and "Social Media".

On the right side, the "Description" pane shows the title "Palestra-Exemplo-Workflow-1", a description "No description has been set yet.", tags "No tags have been added yet.", links "No links have been added yet.", creation date "2020-5-28", and author "Geraldo".

At the bottom, the "Outline" pane shows a tree view of the workflow nodes, and the "Console" pane shows the KNIME Console output, including a welcome message and error messages related to file paths and settings.

**Outline**

# 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 descriptive labels: "Ler arquivo MS-Covid" for the Excel Reader, "Pega apenas dados do Brasil" for the Row Filter, and "Salva em arquivo CSV" for the CSV Writer.

The left sidebar contains the "KNIME Explorer" and "Node Repository" panels. The "KNIME Explorer" shows a tree structure of projects and workflows, including "My KNIME-Hub", "EXAMPLES", and "LOCAL (Local Workspace)". The "Node Repository" lists various nodes categorized by function, such as "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" panel for the selected workflow, which includes fields for "Title", "Description", "Tags", "Links", "Creation Date", and "Author". The "Title" field is filled with "Palestra-Exemplo-Workflow-1".

The bottom panel shows the "Console" window, which displays the KNIME log. The log includes the following text:

```
*** Welcome to KNIME Analytics Platform v4.1.3.v202005121100 ***
*** Copyright by KNIME AG, Zurich, Switzerland ***
Log file is located at: C:\Users\Gerald\AppData\Local\KNIME\workspace\log
WARN Excel Reader: Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer: Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer: Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer: Output file "F:\Github\Palestra-KNIME\Data\Resultado-1.csv" exists and will be overw...
```

# 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

# 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

Filter Criteria | Flow Variables | Memory Policy

☒ Include rows by attribute value  
☐ Exclude rows by attribute value  
☐ Include rows by number  
☐ Exclude rows by number  
☐ Include rows by row ID  
☐ Exclude rows by row ID

Column value matching

Column to test:

☐ filter based on collection elements

Matching criteria

☒ use pattern matching

☐ case sensitive match ☐ contains wild cards  
☐ regular expression

☐ use range checking

lower bound:   
upper bound:

☐ only missing values match

# Tipos de Nós

- Leitura
- Escrita / *Deploy*
- Transformação
- Análise, Aprendizado
- Exploração
- Predição
- Controle

# Tipos de Porta (mais comuns)



Data Table



Flow Variable



Database Connection



Database Query



Image



PMML Model



Tree Ensemble Model



Spark Context



Spark Data



HDFS

# Construindo um Workflow

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

# Onde Estamos?

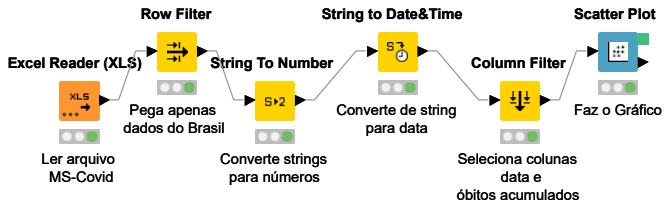
1 O que é o KNIME

2 Workflows

3 Exemplos

4 BIRT – Basic Intelligence Report Tool

# Visualização de Curvas



# Metanodes

**Excel Reader (XLS)**



Ler arquivo  
MS-Covid

**Seleciona Colunas**



Seleciona colunas  
Data e Óbitos Acumulados

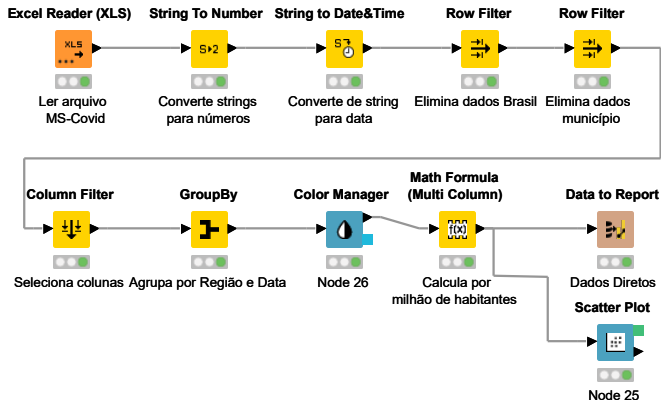
**Scatter Plot**



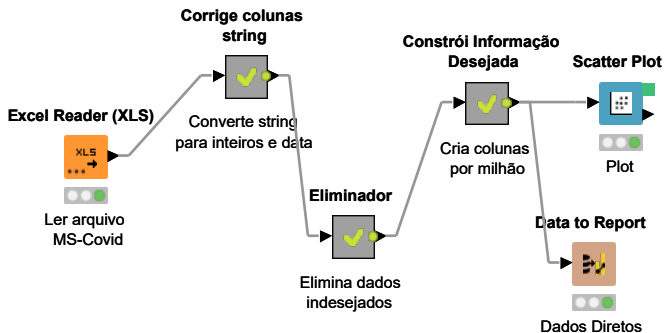
Faz o Gráfico



# Estudando o Covid-19



# Simplificando o Modelo



# Onde Estamos?

1 O que é o KNIME

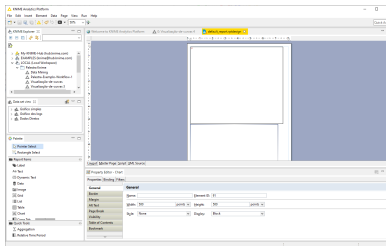
2 Workflows

3 Exemplos

4 BIRT – Basic Intelligence Report Tool

# 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...

# Obrigado!

# Contato



UNIVERSIDADE FEDERAL  
DO RIO DE JANEIRO



INSTITUTO DE MATEMÁTICA  
Universidade Federal do Rio de Janeiro



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



DEPARTAMENTO DE  
CIÊNCIA DA COMPUTAÇÃO  
- UFRJ -



PESC  
Programa de Engenharia  
de Sistemas e Computação



LINE  
Laboratório de Tratamento da  
Informação Não Estruturada

LUDES  
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](#).

