Introdução ao KNIME

Palestras do DCC

Geraldo Xexéo[1,2]

[1]Departamento de Ciências da Computação

[2]Programa de Engenharia de Sistemas e Computação

Palestra Remota para o DCC, Junho 2020

Agenda

Onde Estamos?

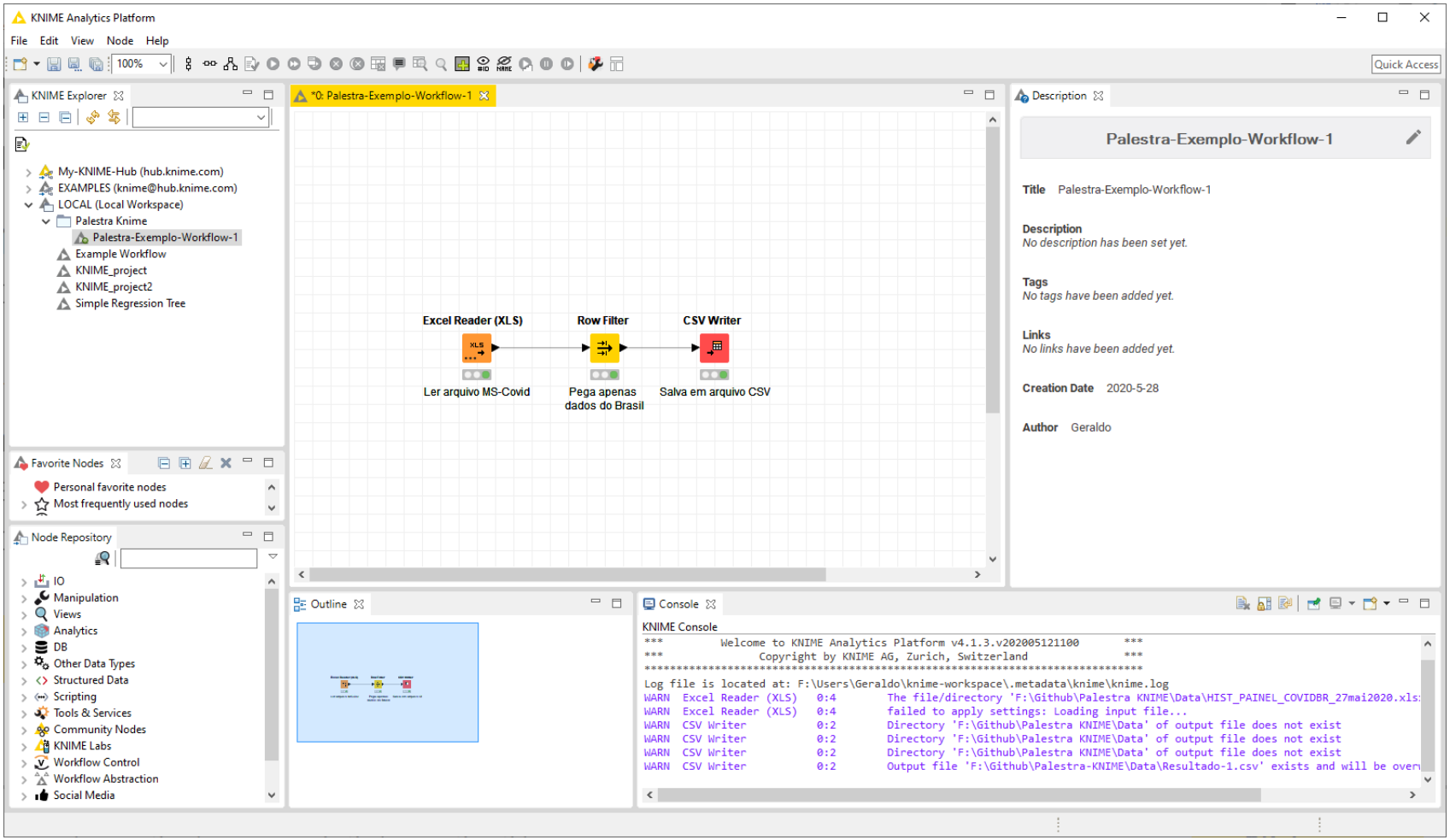
Apresentação do KNIME

- 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

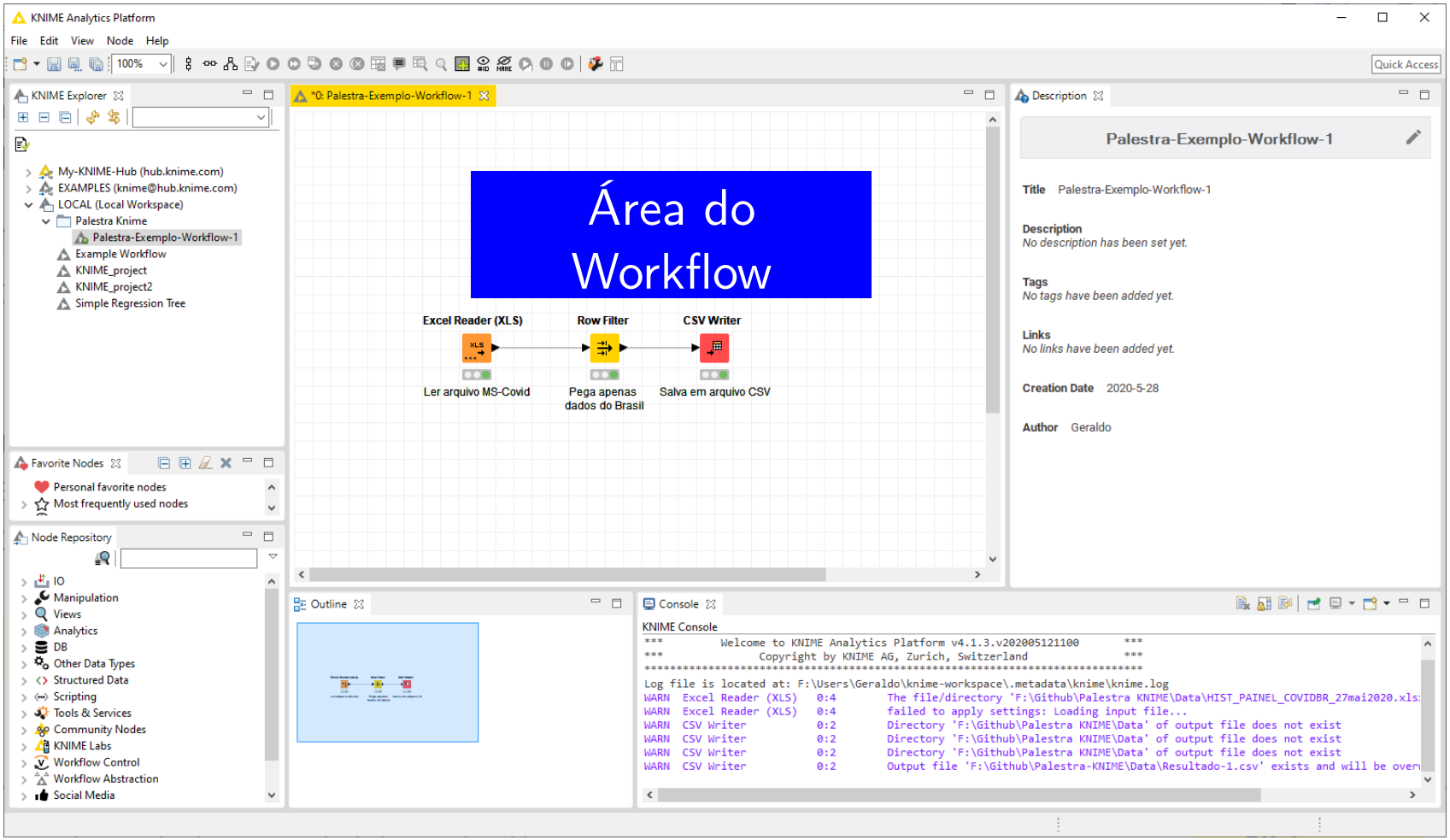
KNIME Analytics Platform

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

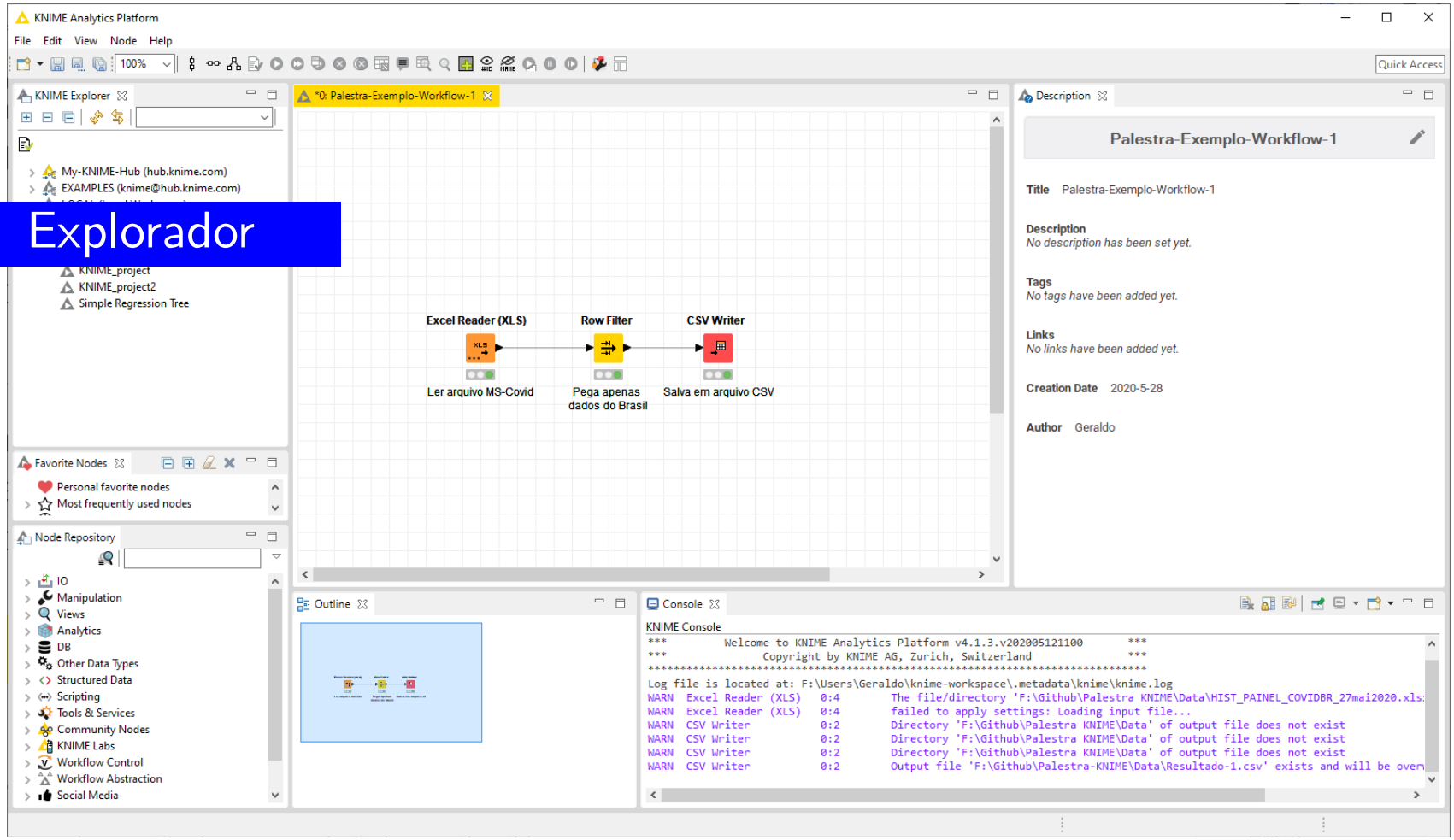
A interface do KNIME



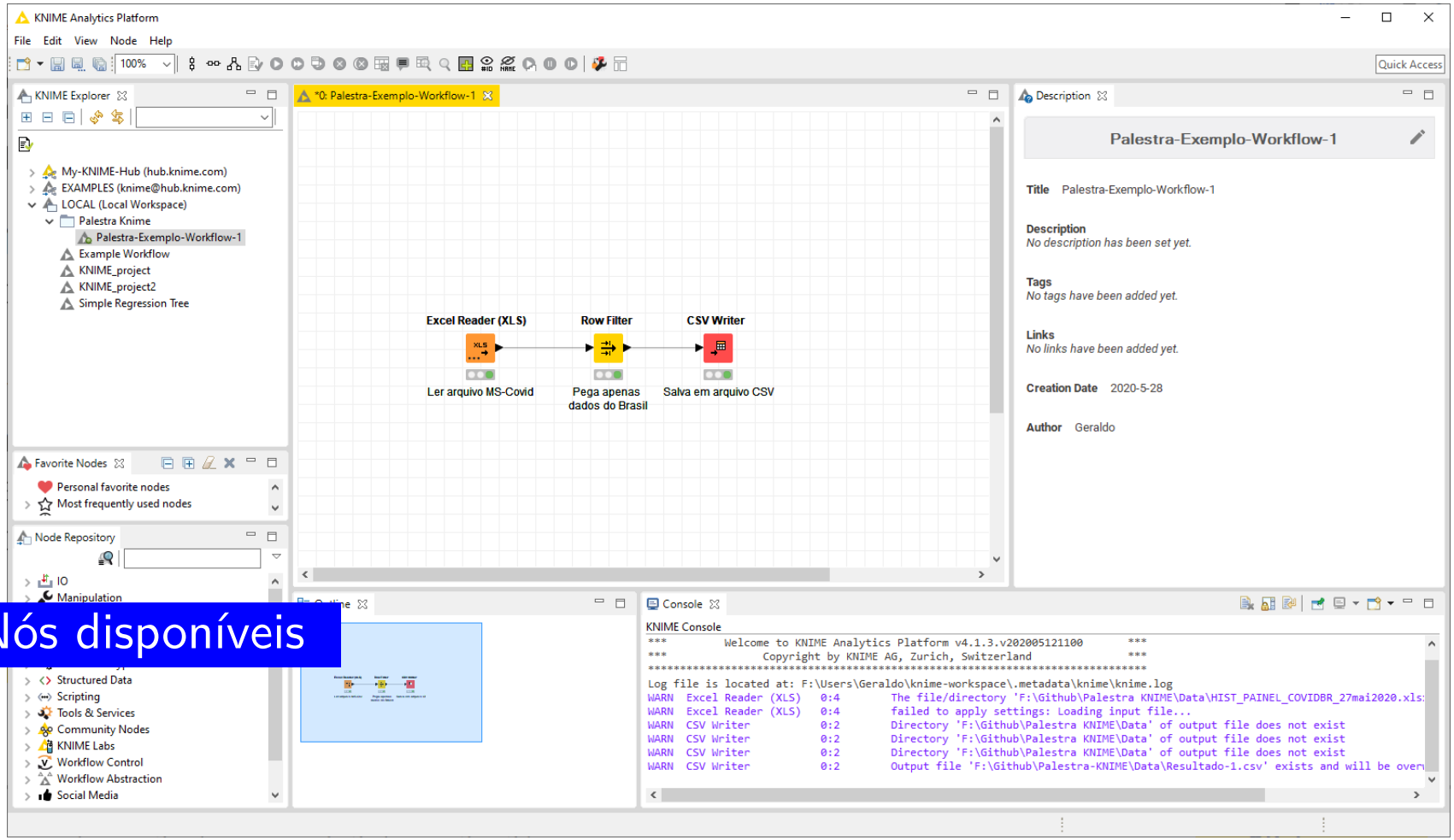
A interface do KNIME



A interface do KNIME

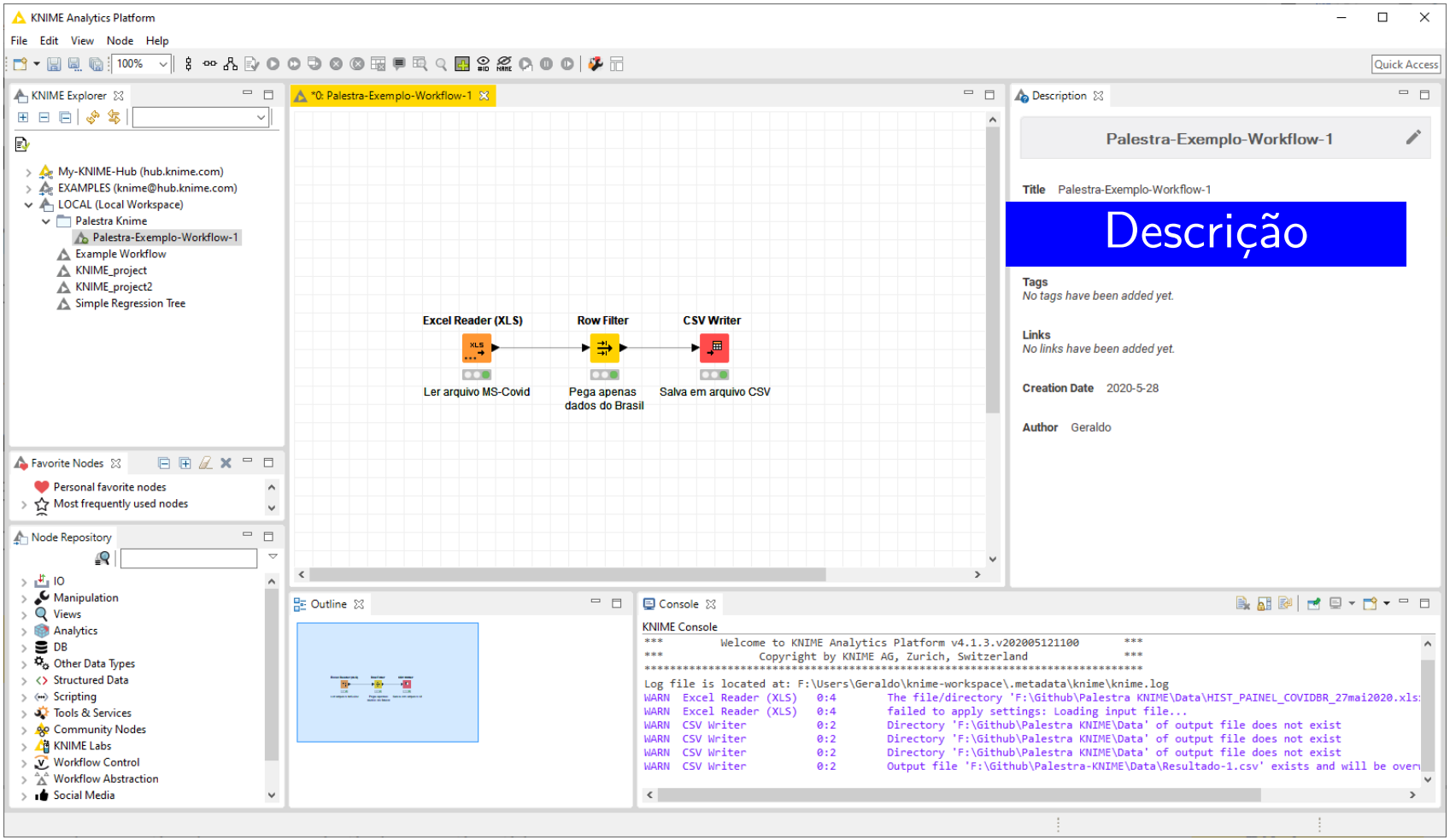


A interface do KNIME

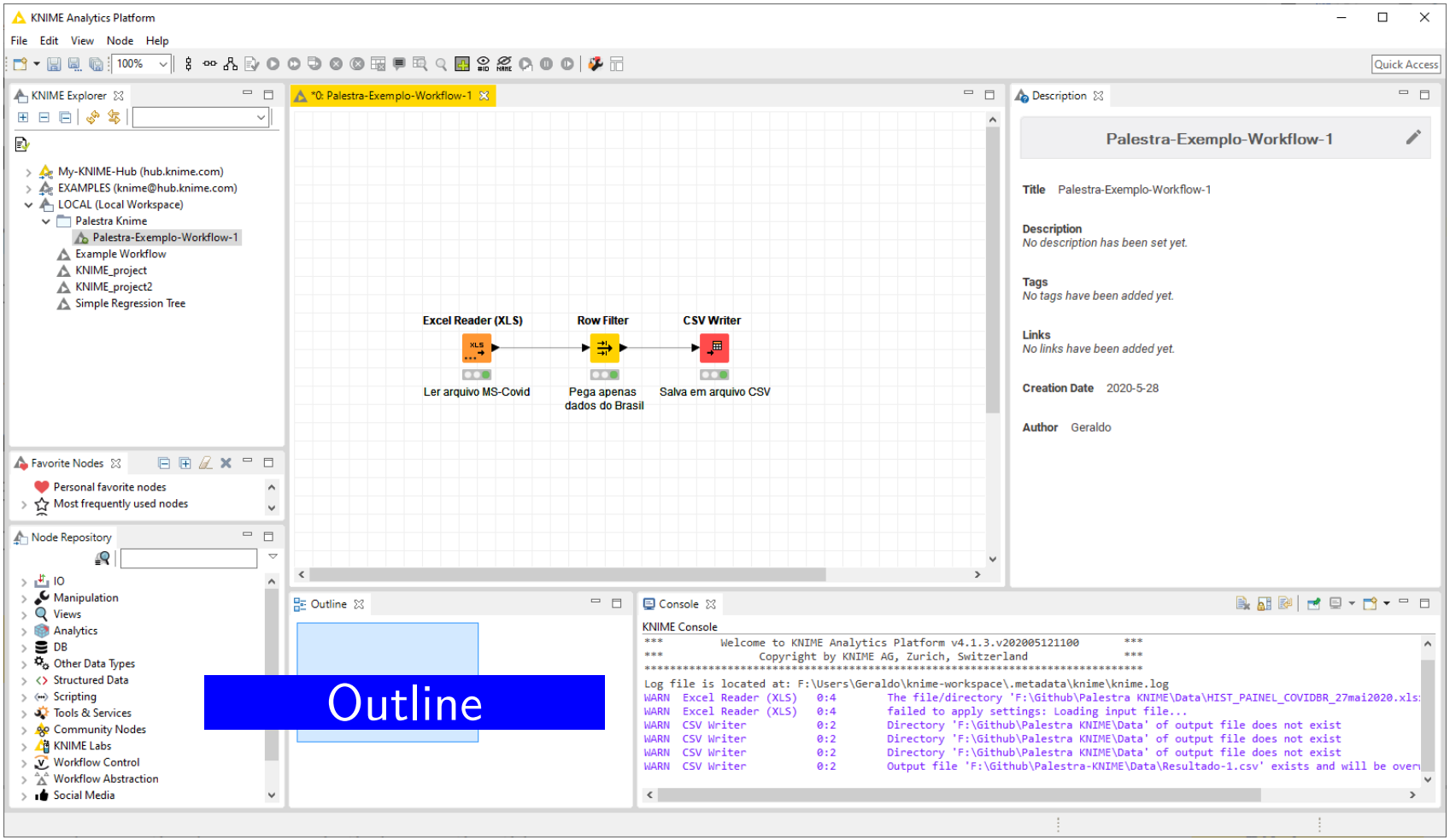


Nós disponíveis

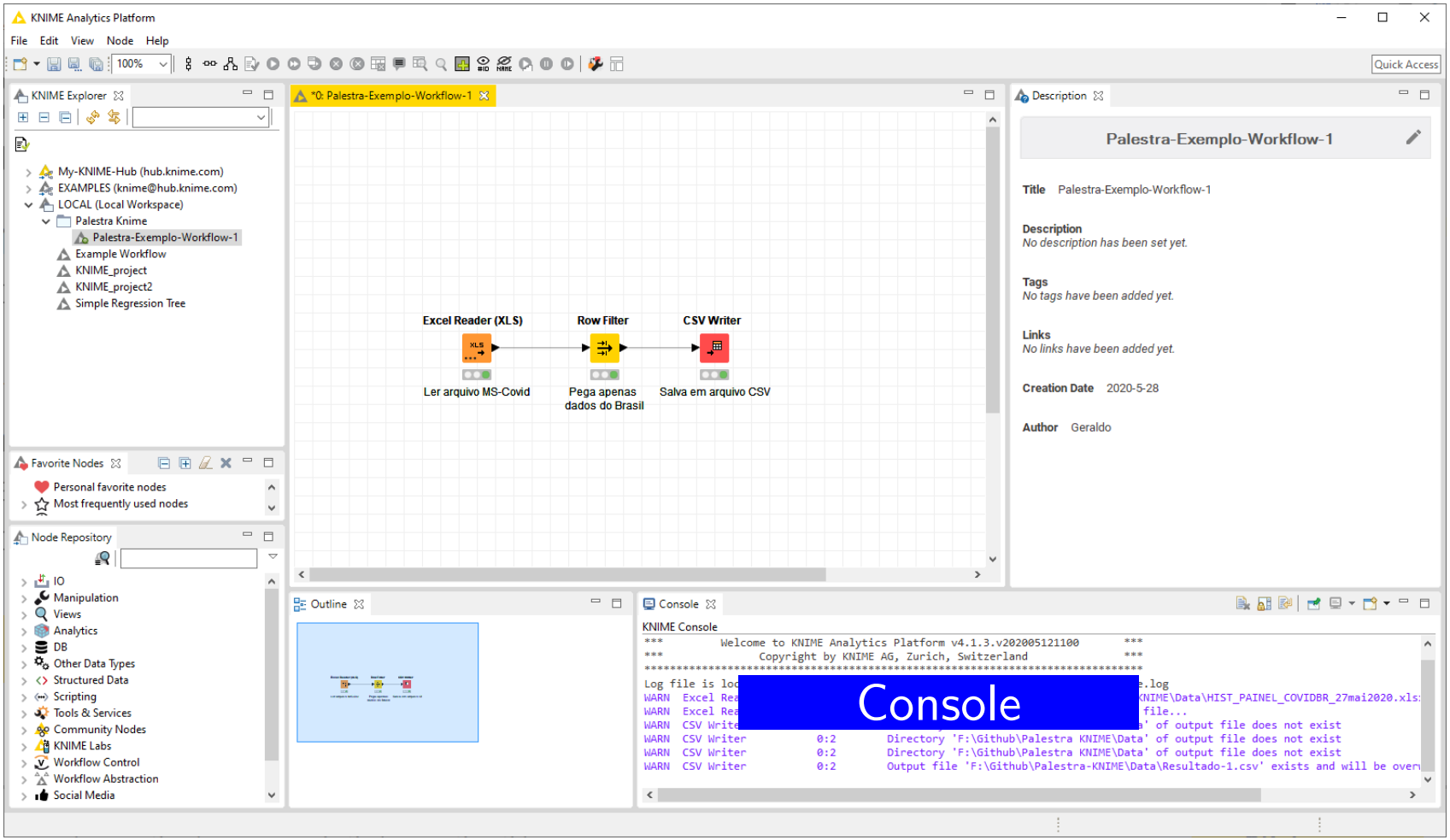
A interface do KNIME



A interface do KNIME

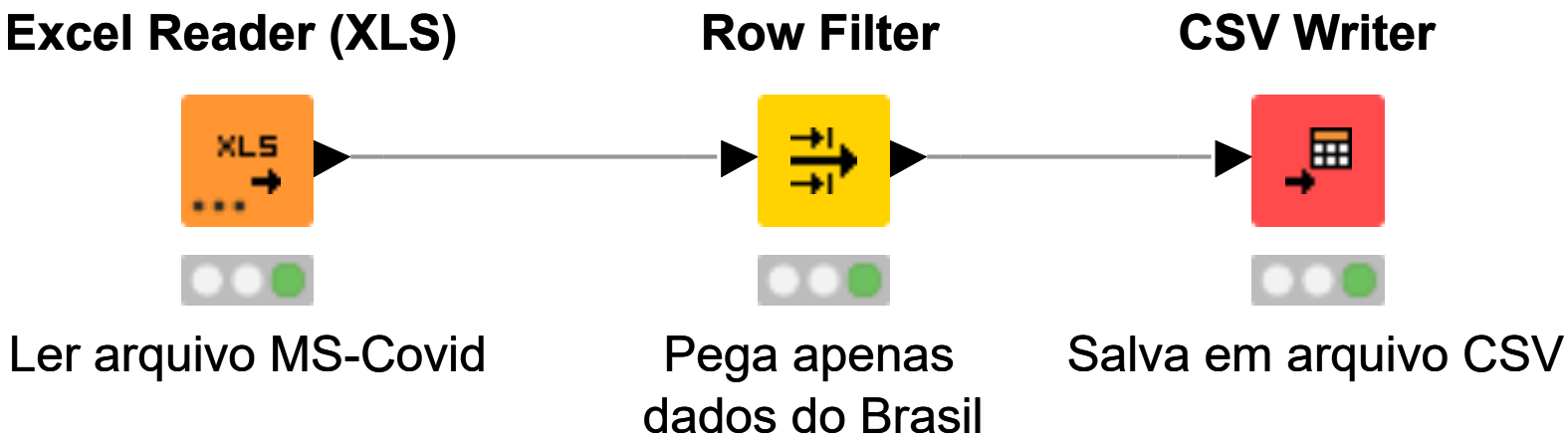


A interface do KNIME



O que é um workflow

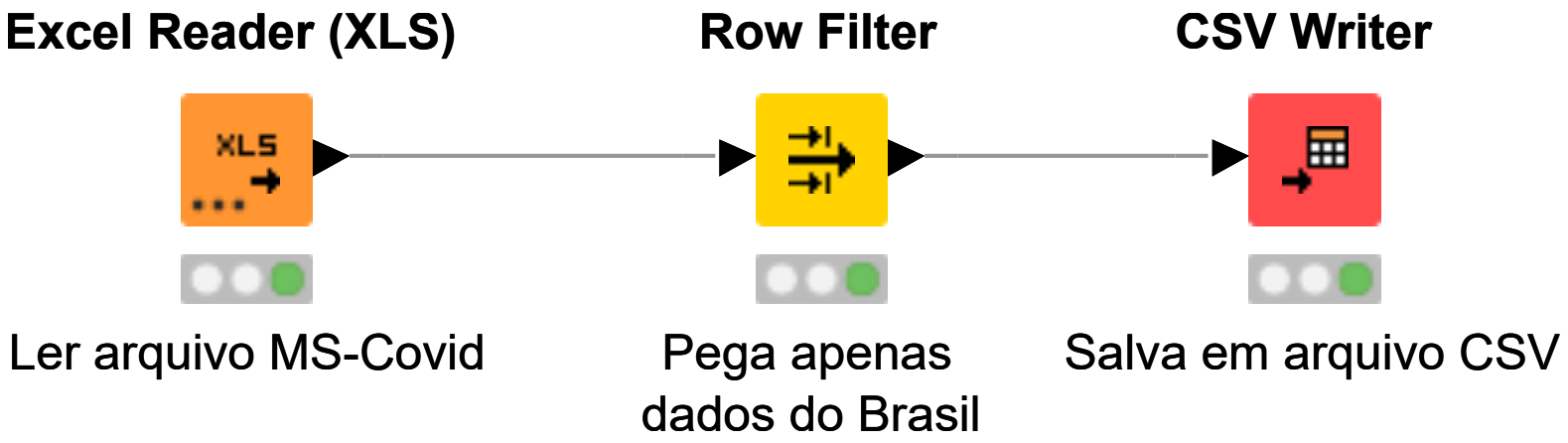
Um workflow, ou fluxo de trabalho, é uma forma de descrever uma cadeia de processamento por meio de um grafo onde nós representam processos e arestas representam o fluxo de dados entre esses processos.



Onde Estamos?

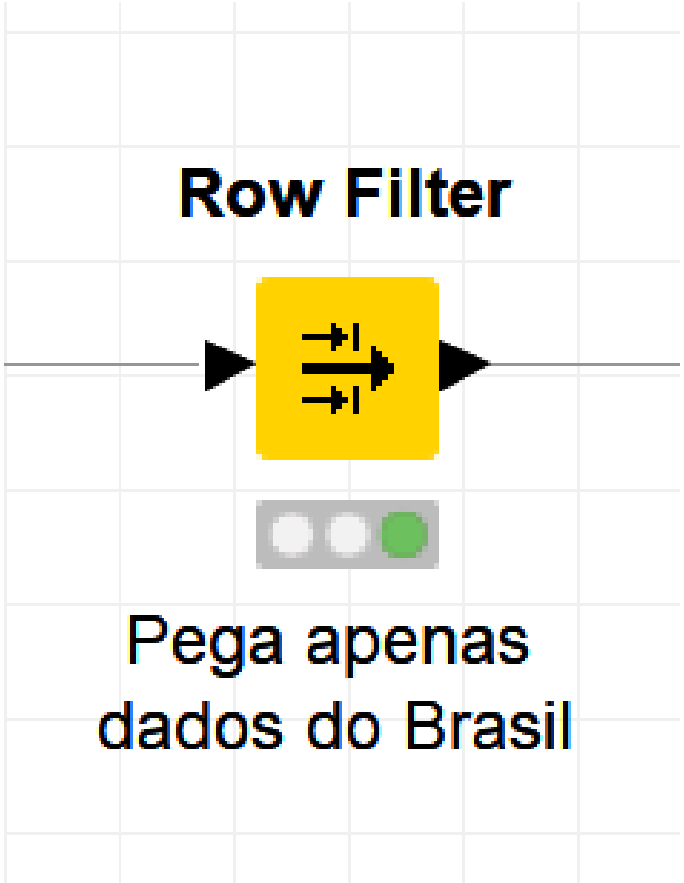
Entendendo um Workflow

- 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



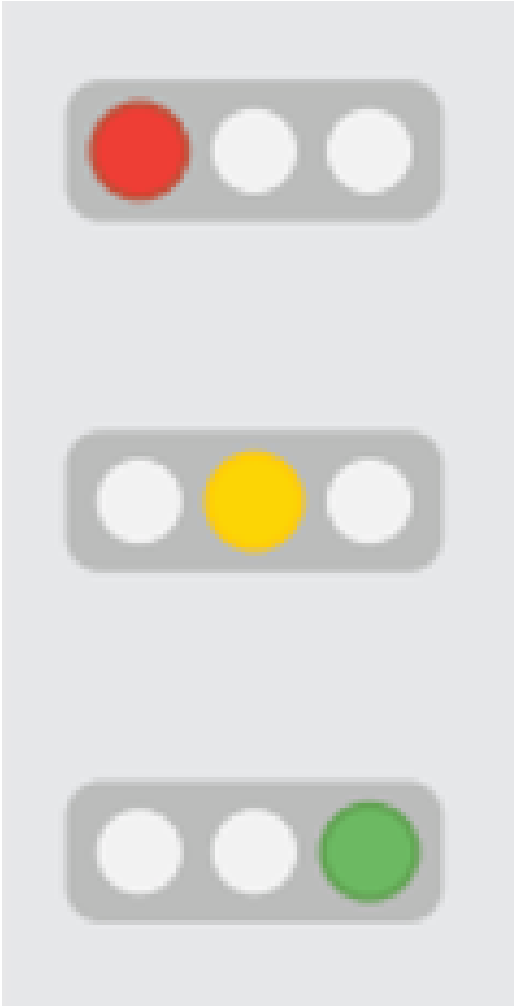
Um nó básico

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

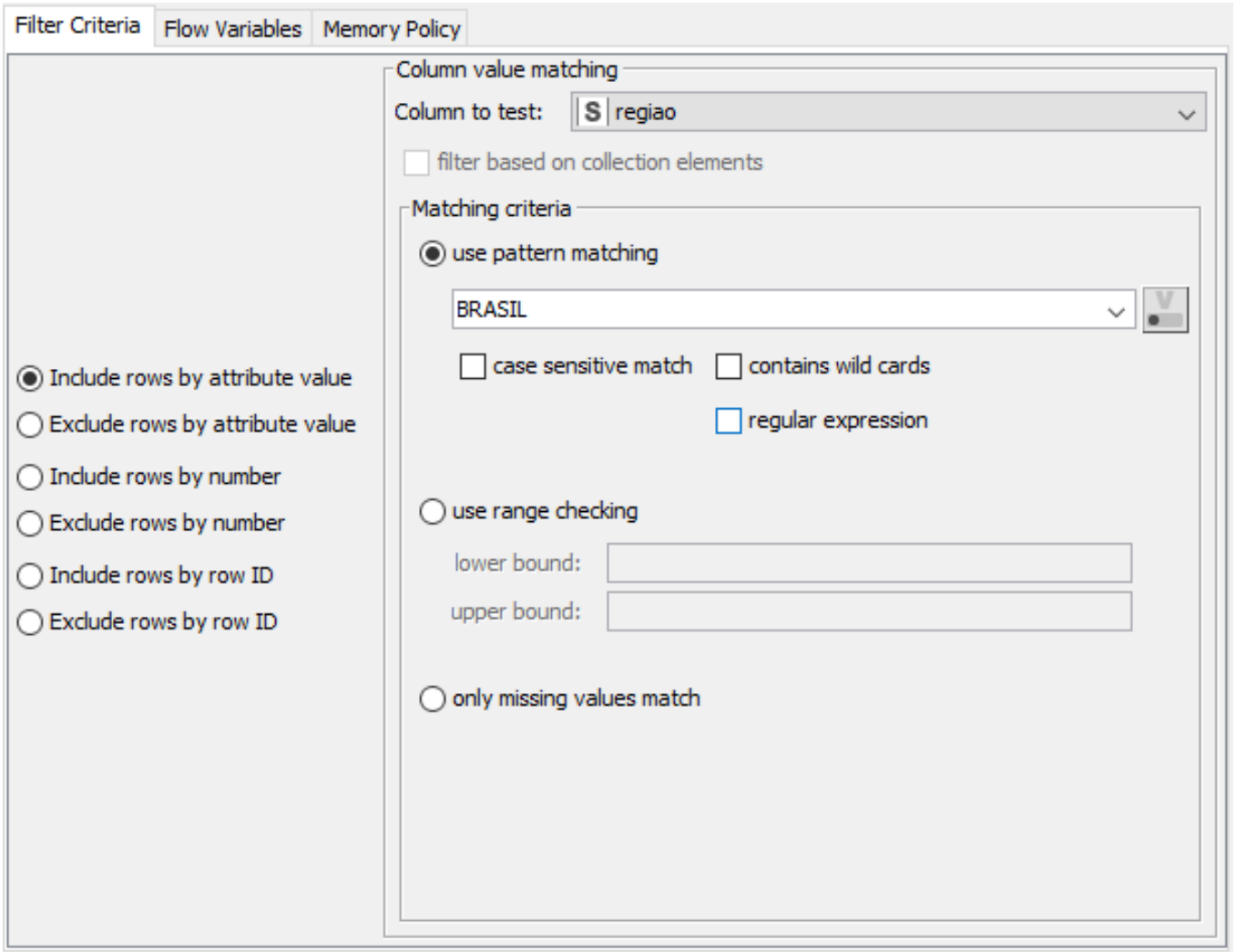


Status do Nó

- Não configurado
- Pronto para executar
- Executado



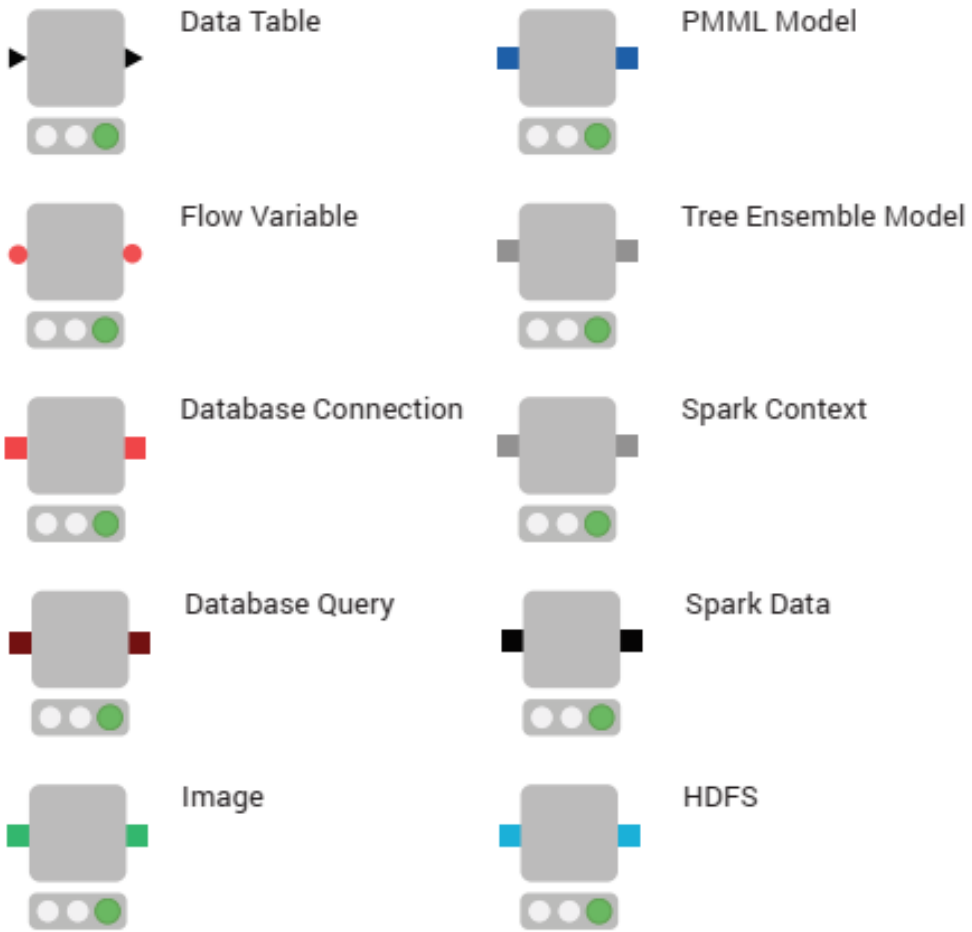
Configuração de um Nó



Tipos de Nós

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

Tipos de Porta (mais comuns)

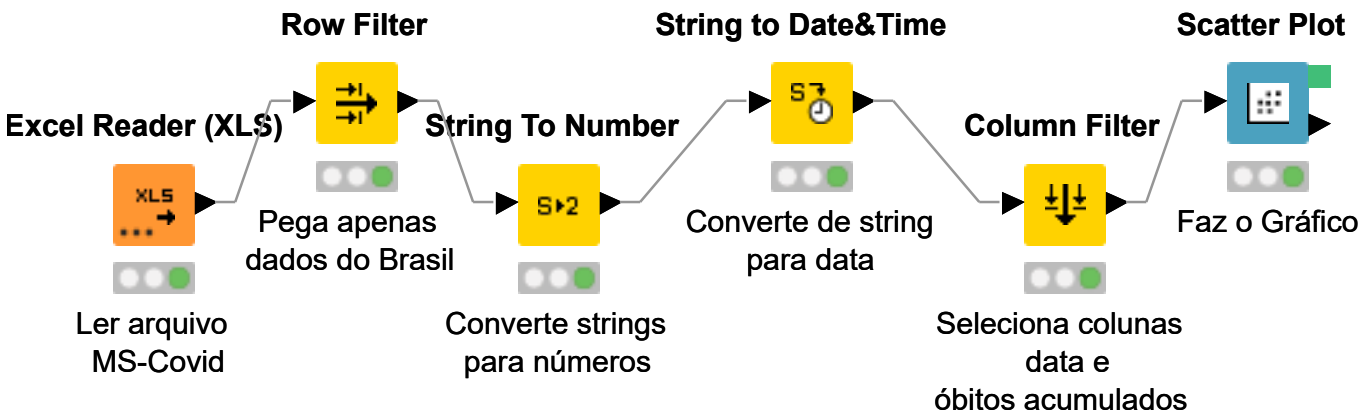


Construindo um Workflow

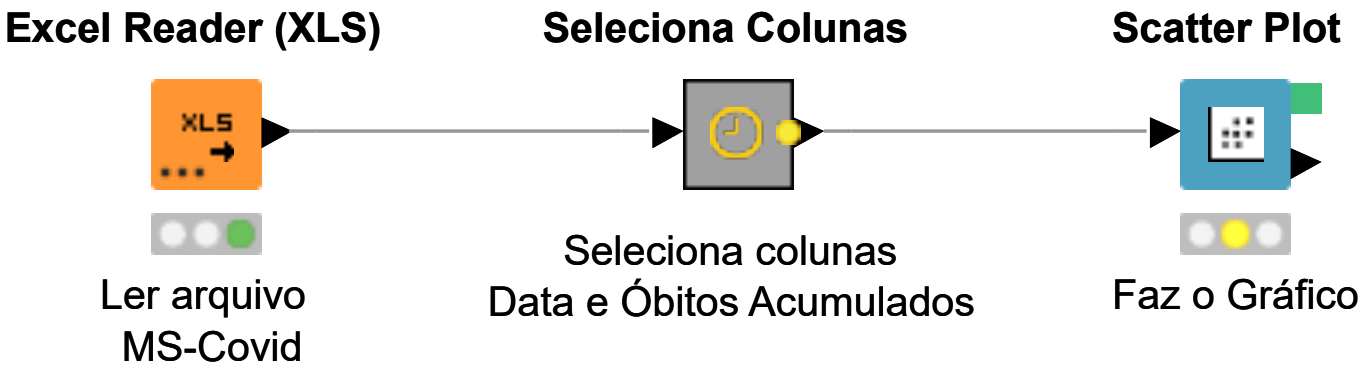
- Coloque nós no mapa
- Configure os nós
- Execute
- Repita até alcançar o desejado

Onde Estamos?

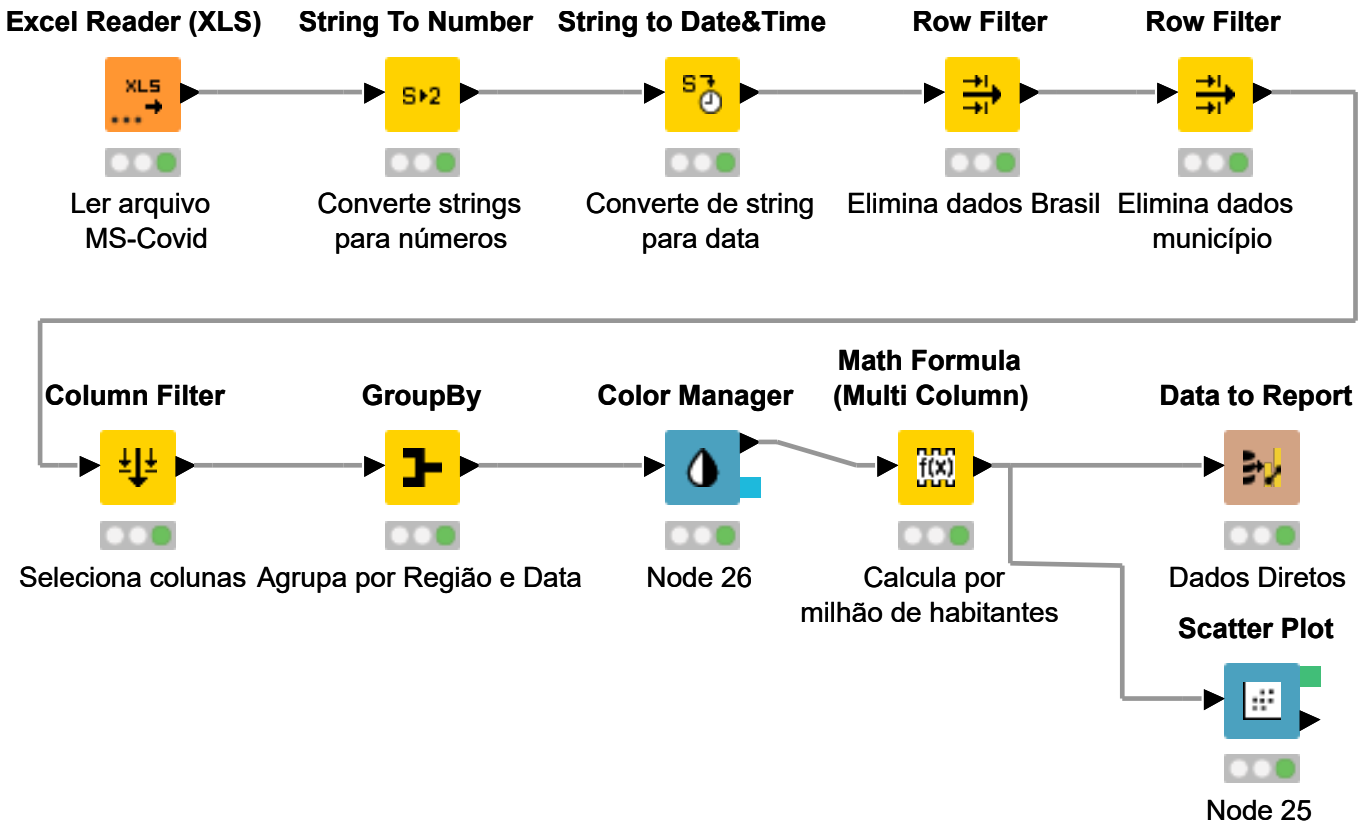
Visualização de Curvas



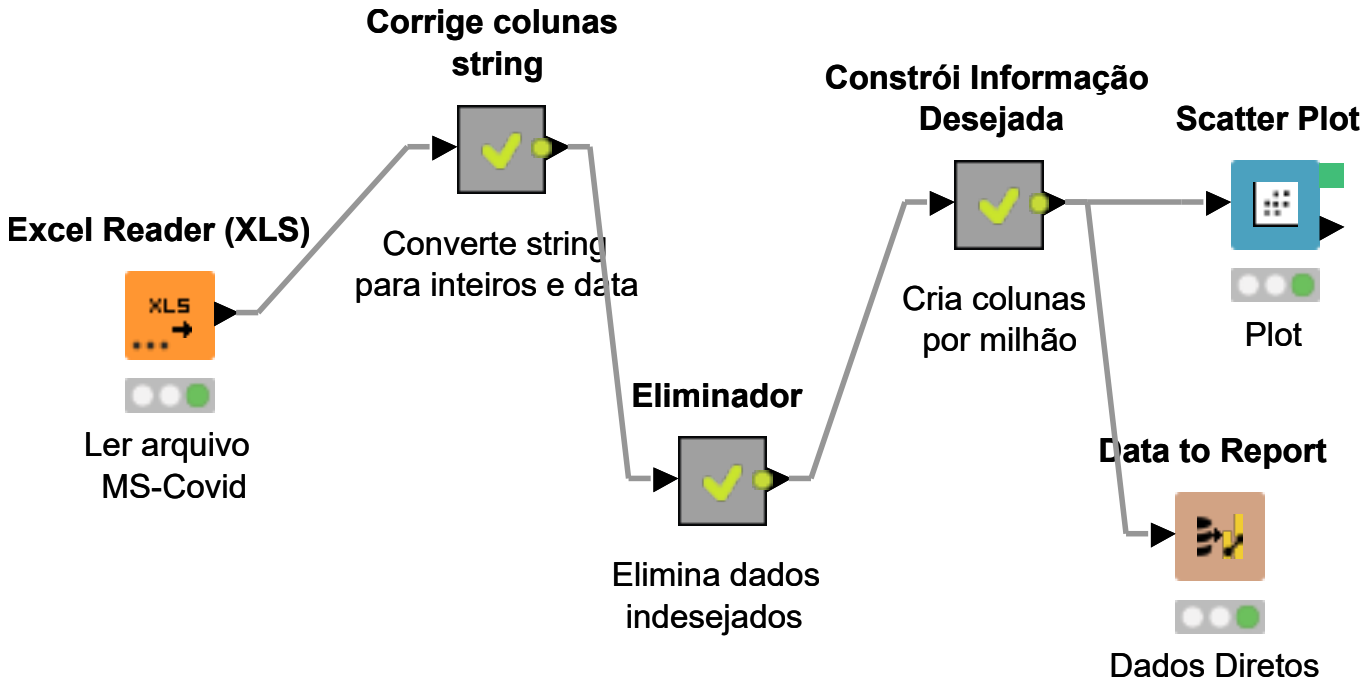
Metanodes


Excel Reader (XLS)
Seleciona Colunas
Scatter Plot
XLS
Ler arquivo
MS-Covid
Seleciona colunas
Data e Óbitos Acumulados
Faz o Gráfico

Estudando o Covid-19



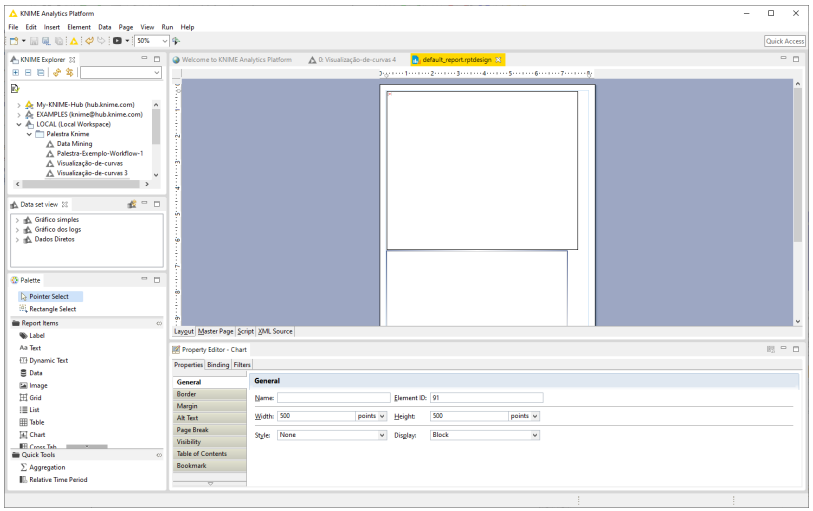
Simplificando o Modelo



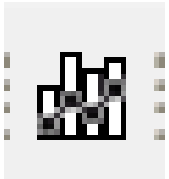
Onde Estamos?

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 págica
- defina a master page
- usos mais avançados com scripts...

Obrigado!

Contato

Geraldo Xexéo
xexeo@cos.ufrj.br
gxexeo@gmail.com