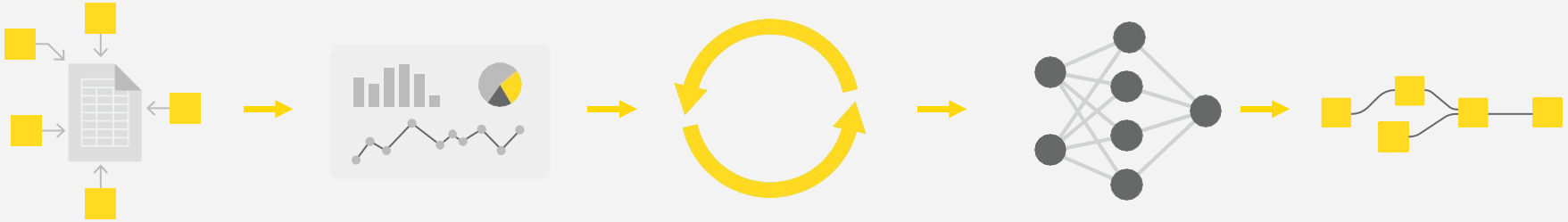
Data Exporting & Reporting

Who is this course for?

Data Analyst	Data Scientist	Data Engineer
Data acquisition, cleaning, analysis, visualizations, descriptive statistics, reporting, dashboards.	Data pre-processing, training machine learning and statistics algorithms, modeling, predicting.	Integrating various data sources, building data pipelines (ETL, ELT), databases, data lakes, data warehouses, file systems, and/or data mart maintenance, monitoring and testing.

Data analytics within an organization



Data Collection

Gathering data from different sources

Data Visualization

Data visualization and statistics

Data Preprocessing

Cleaning the data; checking quality

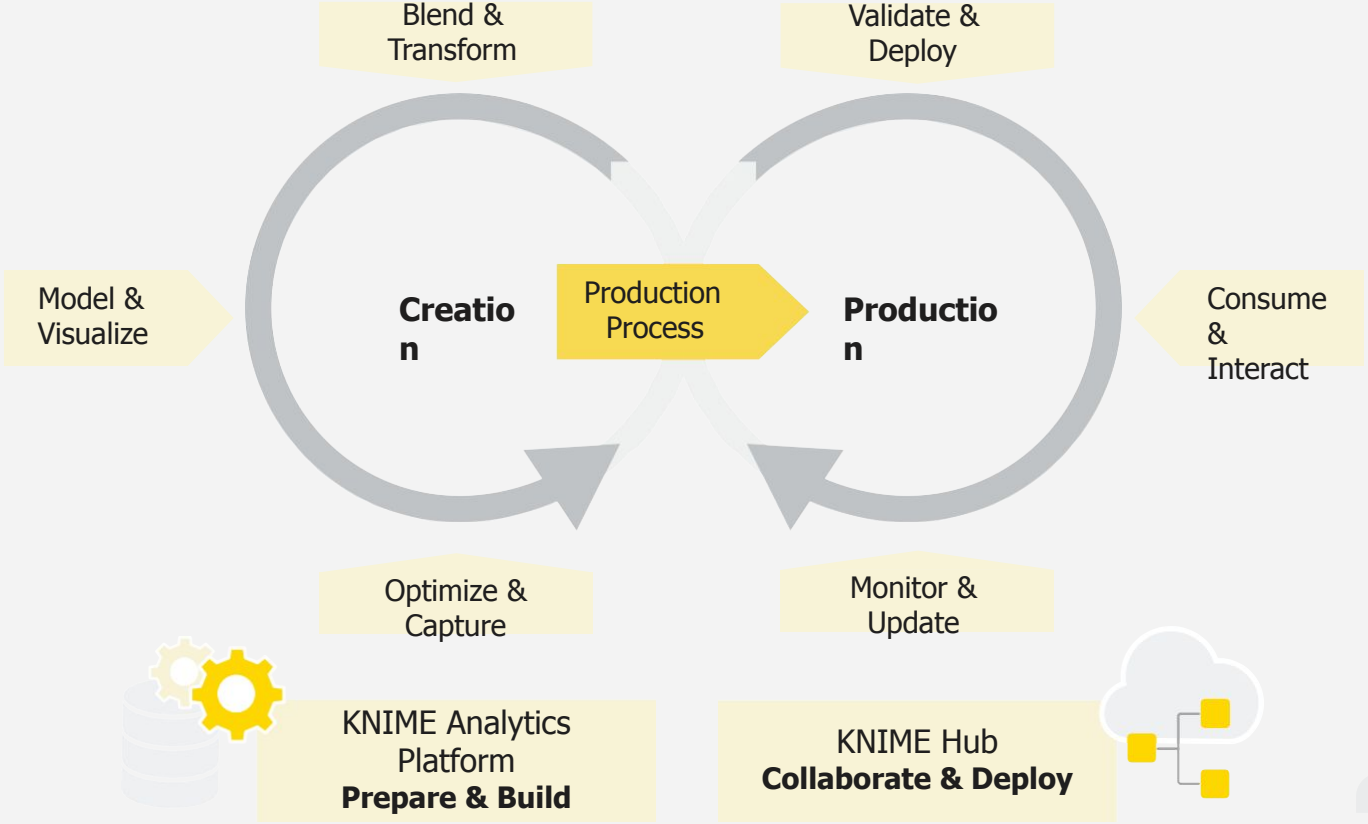
Machine Learning

Modeling and evaluation

Deployment for Users

Production model

The Data Area Life Cycle



Exporting Data & Deployment

Exporting Data

After an analysis is completed, what next?

- Write results to a file
- Create/update a database
- Save the model for use elsewhere
- Generate a rich report
- Deploy via KNIME WebPortal
- Deploy via workflow as RESTful web service

Input/Output in Deployment

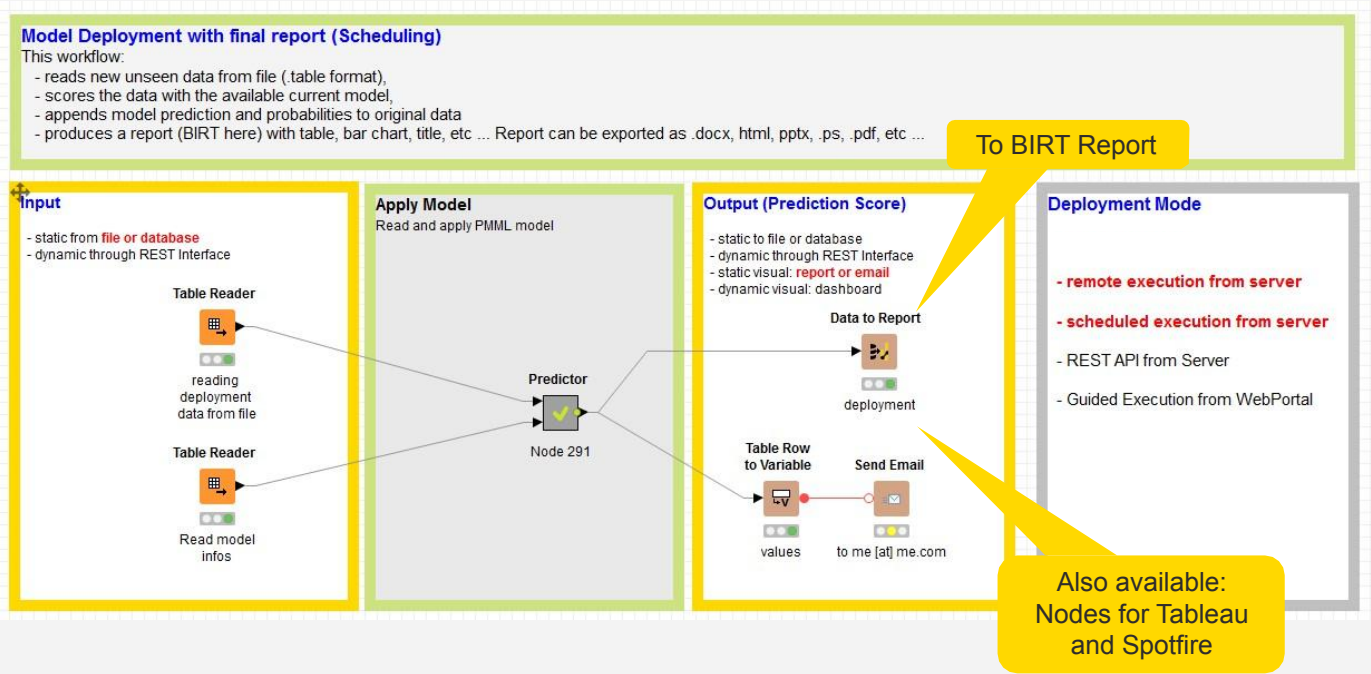
Input

- File (CSV, Table, XLS, ...)
- Database
- JSON for REST API

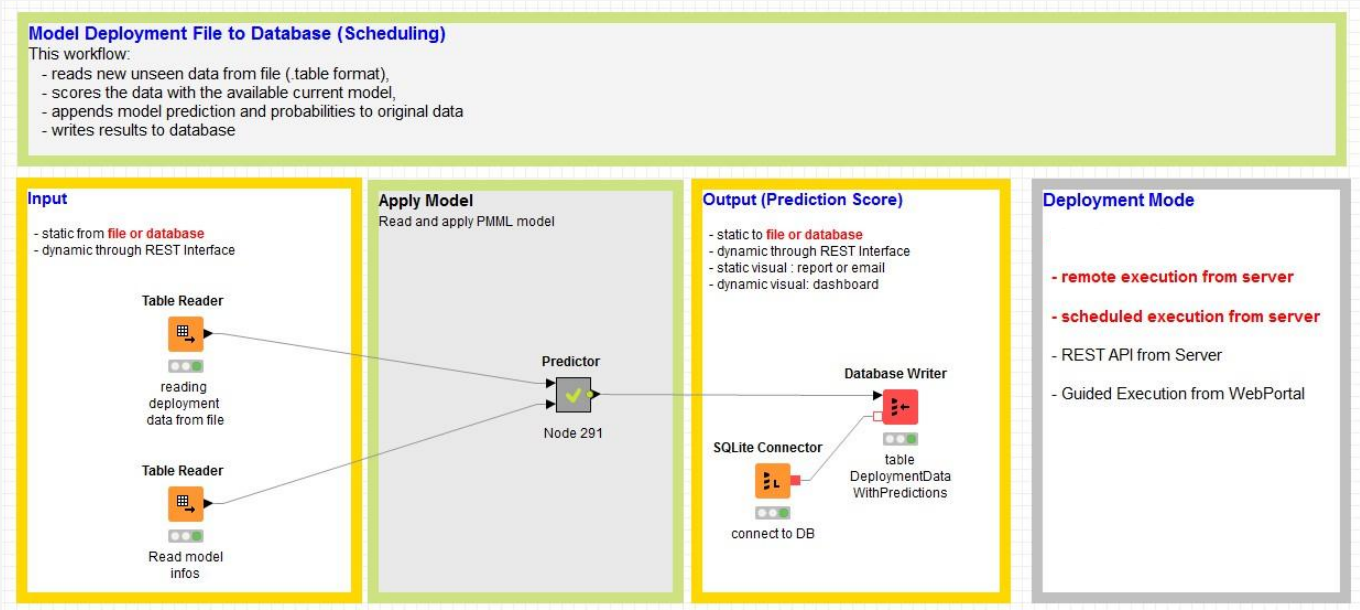
Output

- Report (BIRT, Tableau, Spotfire, PowerBI)
- Email
- File (CSV, Table, XLS, ...)
- WebPortal

To Report / Email



To File / Database

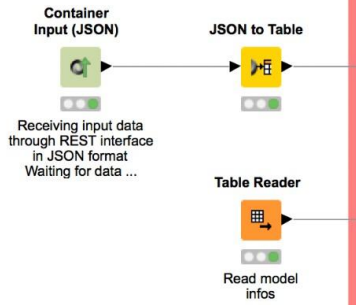


REST API (Available on KNIME Server)

Model Deployment as REST API
This workflow:
- receives new unseen data via REST interface (JSON format),
- scores the data with the available current model,
- appends model prediction and probabilities to original data,
- makes results available at the output REST interface.

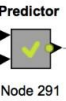
Input

- Static data from file or database
- **Dynamic through REST Interface**



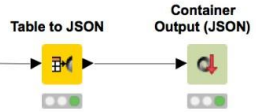
Apply Model

- **PMML predictor**
- **Dedicated predictor**
- SQL model recoding
- Executable JAR model



Output

- Static to file or database
- **Dynamic through REST Interface**
- Static visual: report or email
- Dynamic visual: dashboard
- External application update



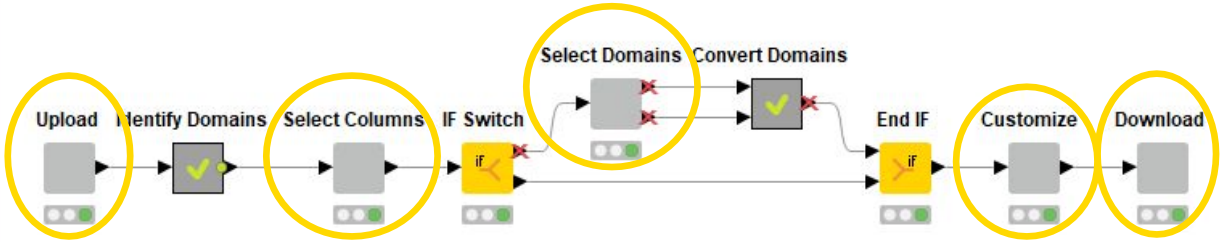
Deployment Mode

- Remote execution from server
- Scheduled execution from server
- **External application**
- **REST API from Server**
- Guided Execution from WebPortal

To Dashboard on WebPortal

The Process Step by Step

1. Upload your data / Select one of the available datasets
2. Select the columns to visualize (maximum 3)
3. Convert the domain of the columns (OPTIONAL)
4. Customize the visualizations interactively
5. Download the images of the customized charts



Step 1
Upload File

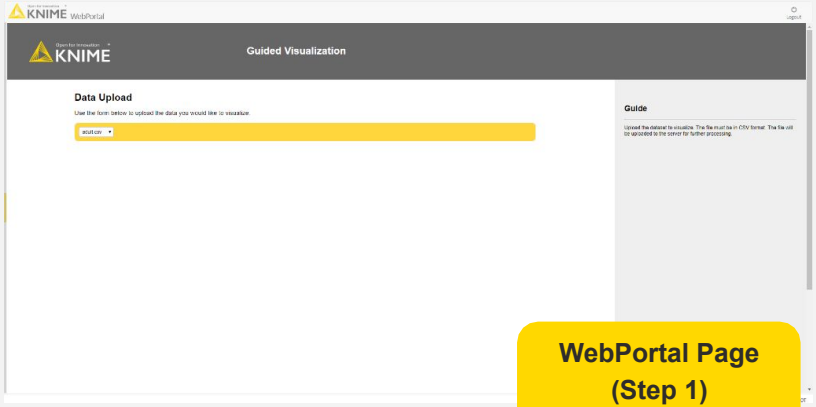
Step 2
Select Columns

Step 3
Customize Column Domains

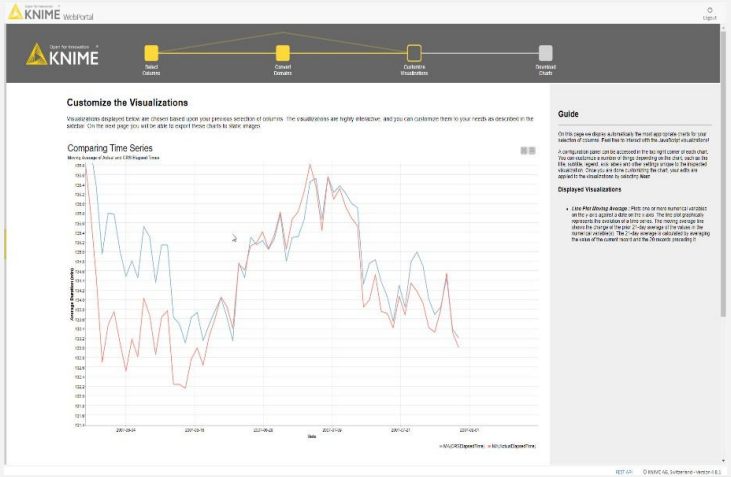
Step 4
Interactive View

Step 5
Download Image

Workflow on KNIME WebPortal



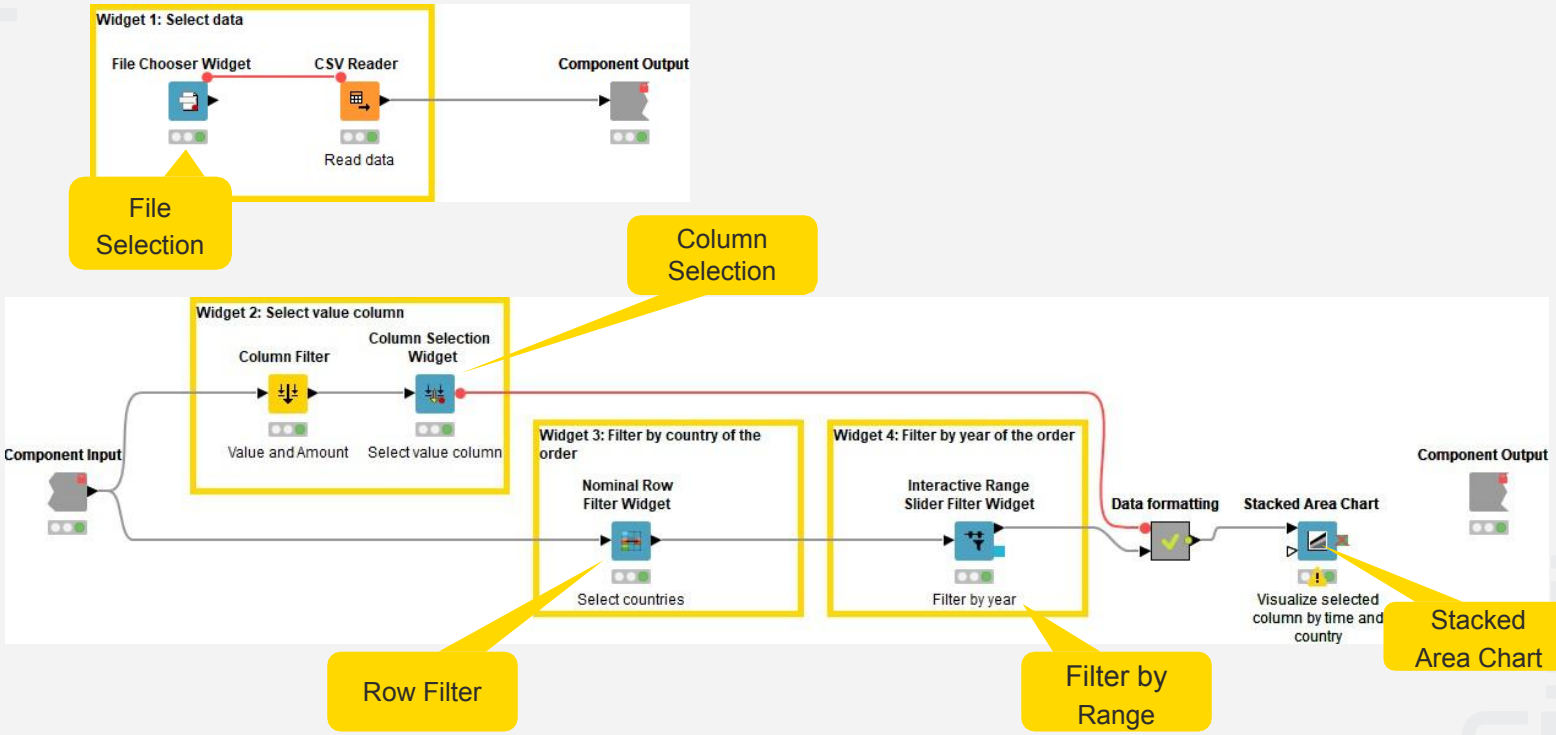
WebPortal Page (Step 1)
Upload File



WebPortal Page (Step 4)
Interactive View

Available in KNIME Server

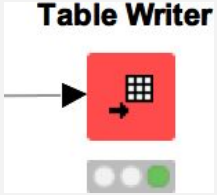
Components to Produce Dashboard on Web Page

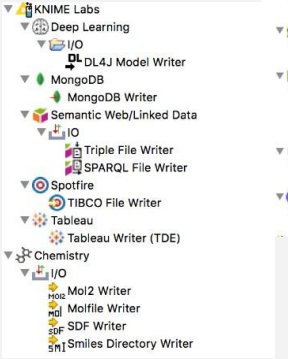


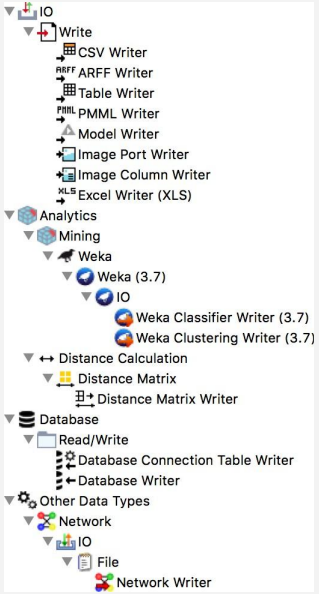
Data Export Nodes

Typically characterized by:

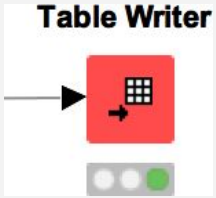
- Magenta color
- 1 input port, no output ports
- Create file on file system or write to database

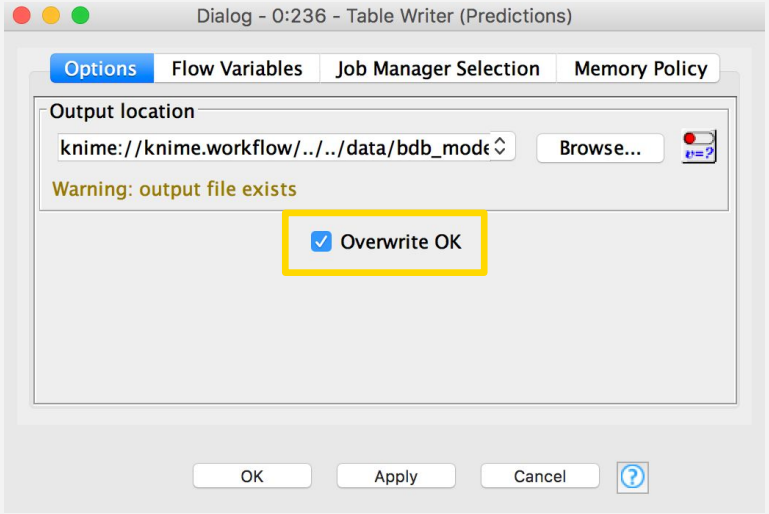




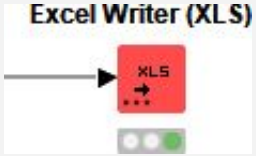


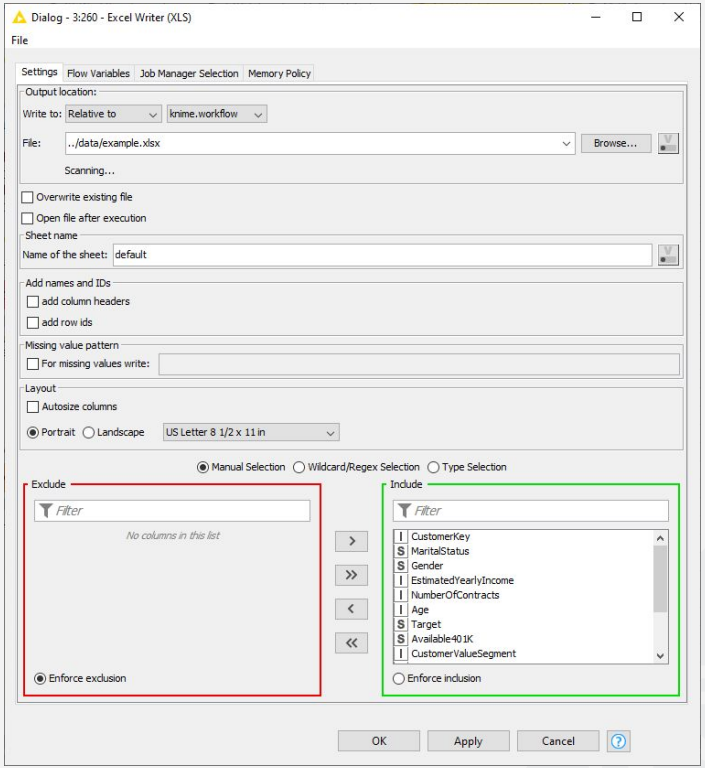
New Node: Table Writer



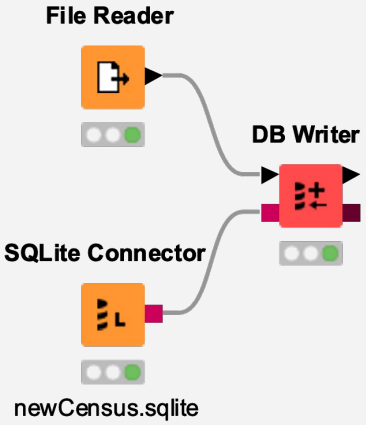


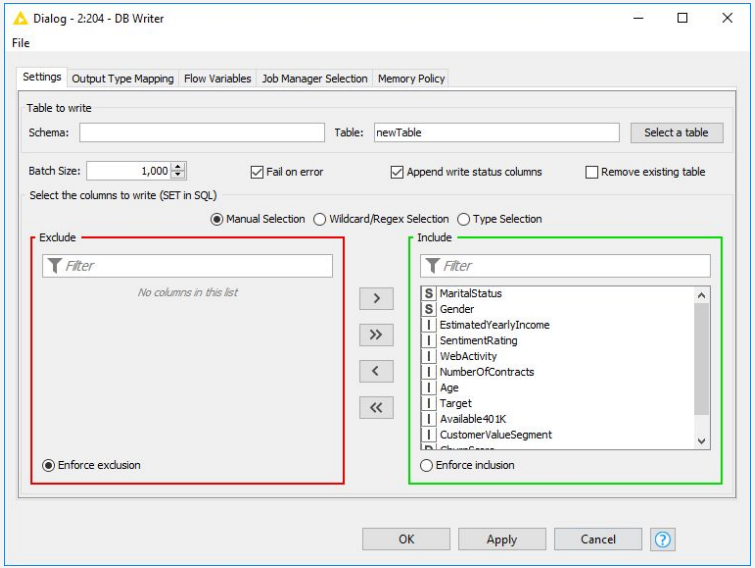
New Node: XLS Writer





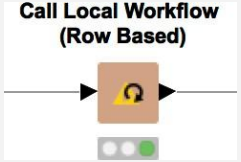
New Node: Database Writer

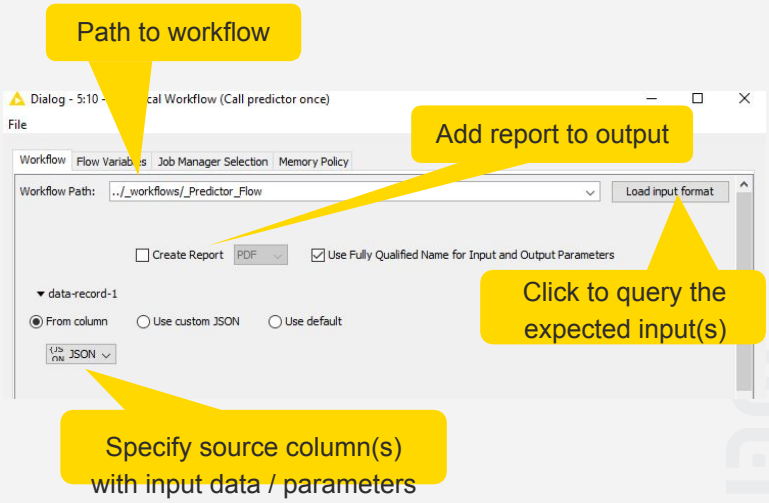




Automation: Call Local Workflow

- Use Call Local Workflow node to send data and parameters to other workflows and trigger execution
  - Send results back to caller-workflow
  - Include report from called workflow
- Create modular workflows
  - E.g. separate workflows for ETL and prediction
- Alternative: Call Remote Workflow
  - Trigger execution of workflows on KNIME Server via REST API

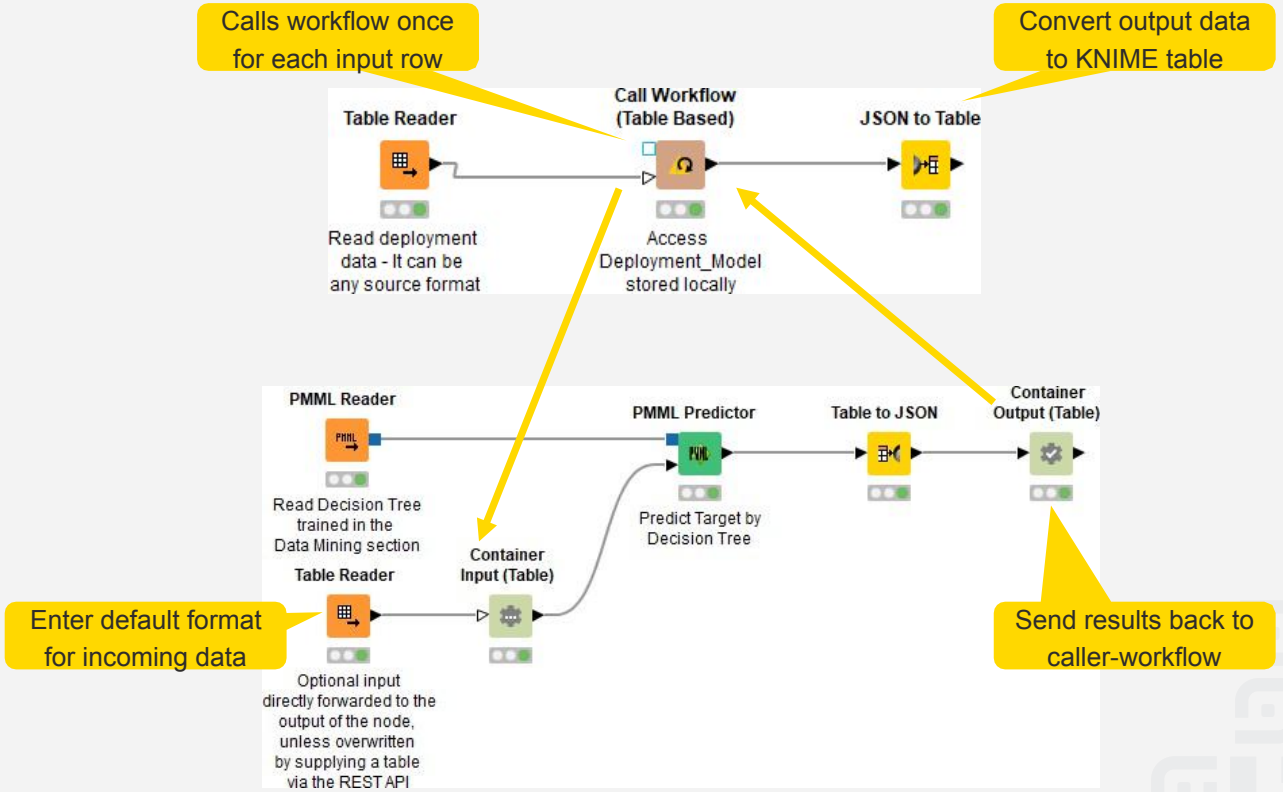




Automation: Call Local Workflow

ETL

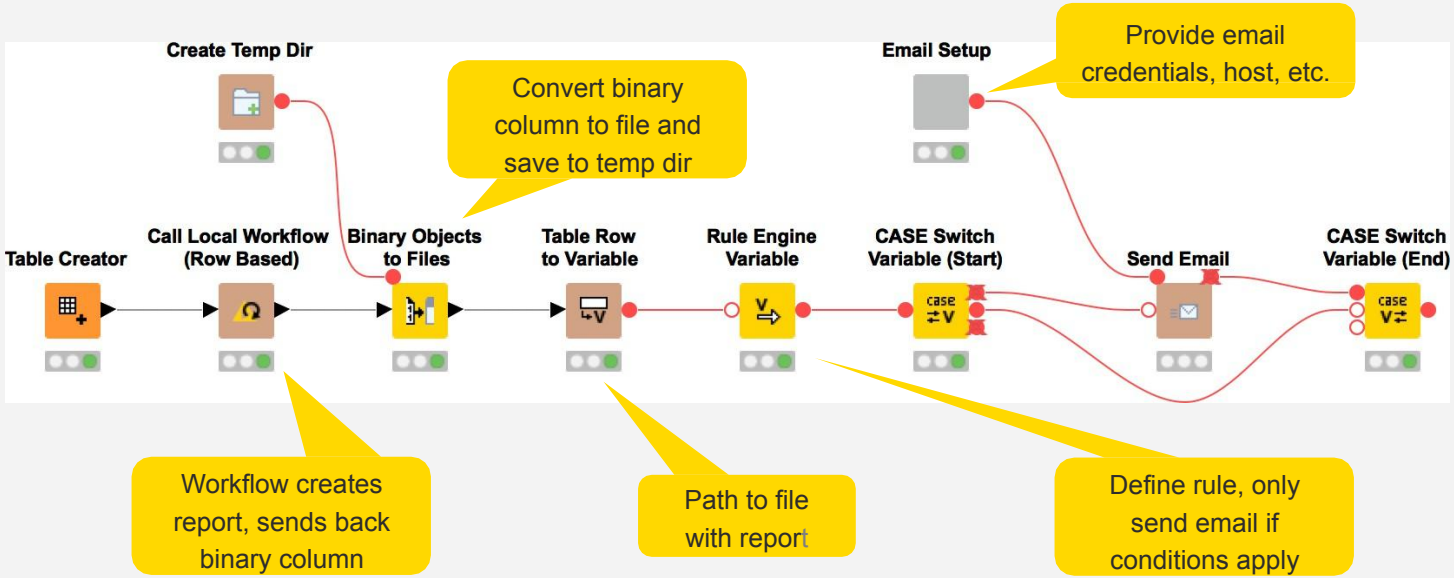
Prediction



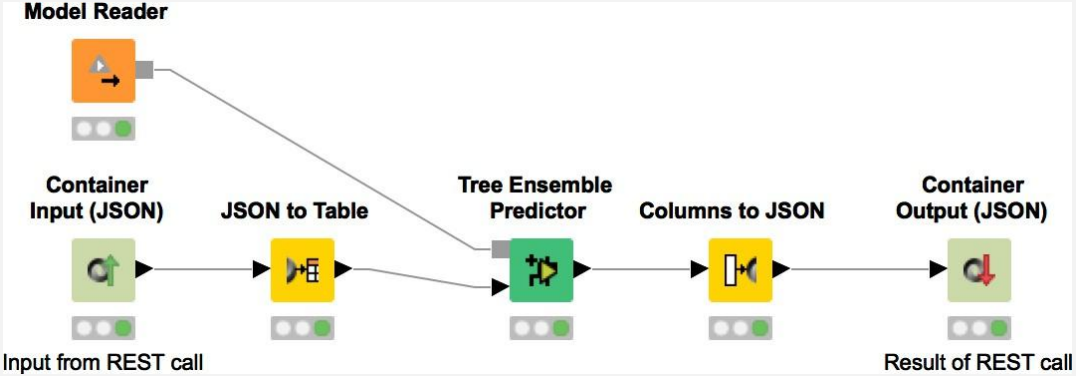
Use Call Local Workflow to Send Conditional Emails with Report

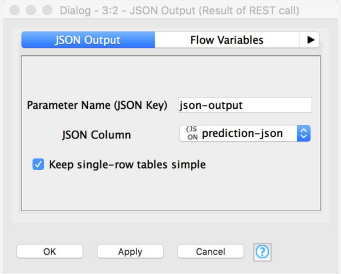
Sometimes, report should be sent under specific circumstances

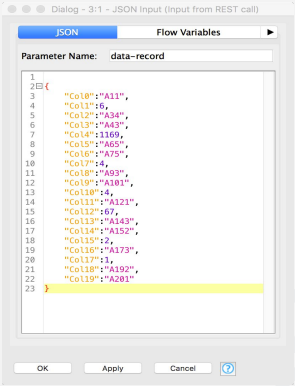
- E.g. if some KPI is below threshold



KNIME Server as a REST Resource



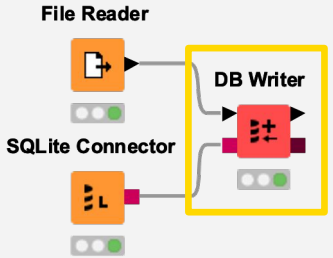


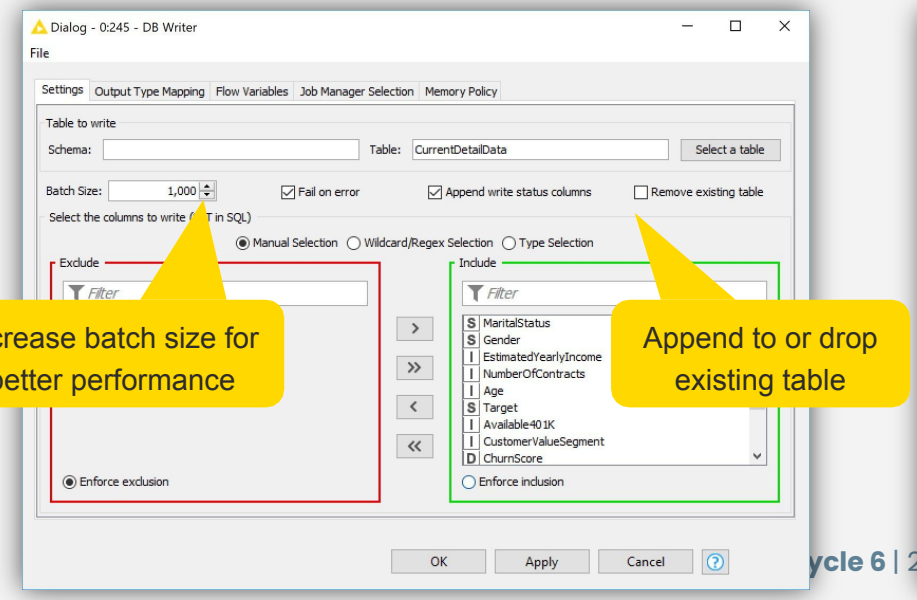


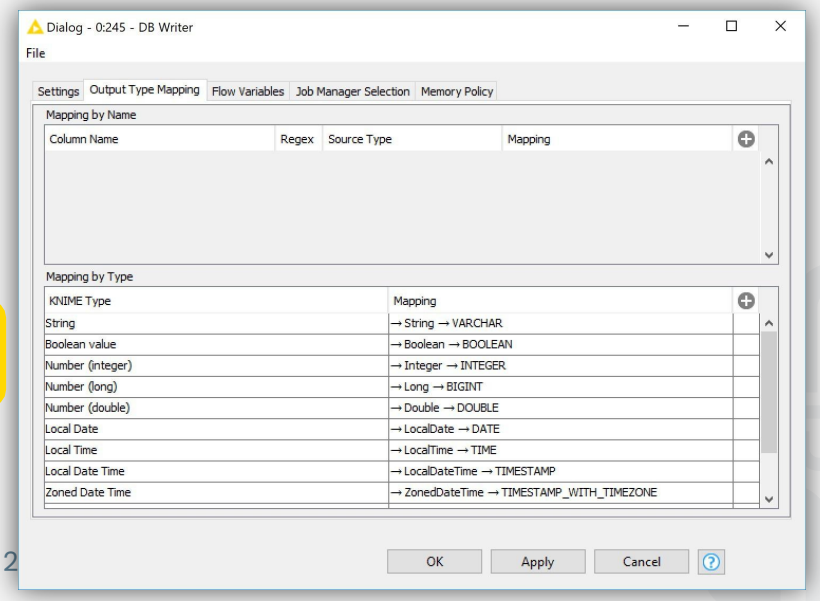
https://www.knime.org/blog/giving-the-knime-server-a-rest

DB Writer

- Writes data from a KNIME data table **directly** into a database table

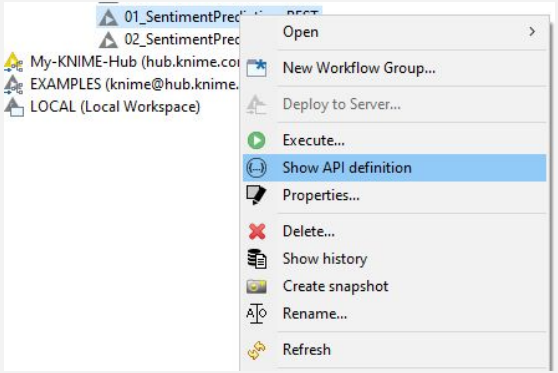


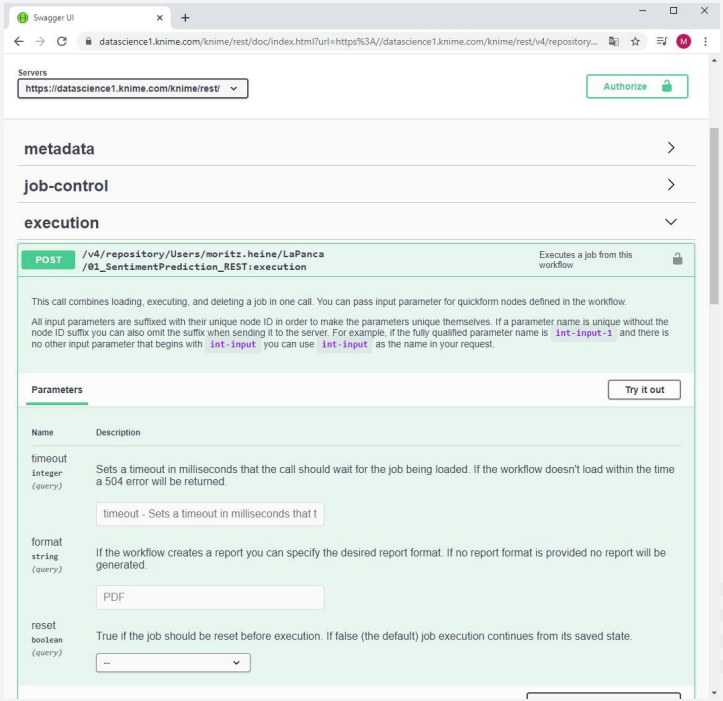




KNIME Server as a REST resource

- Use Swagger, SOAPUI or Chrome extension Postman to explore the HTTP requests and test them

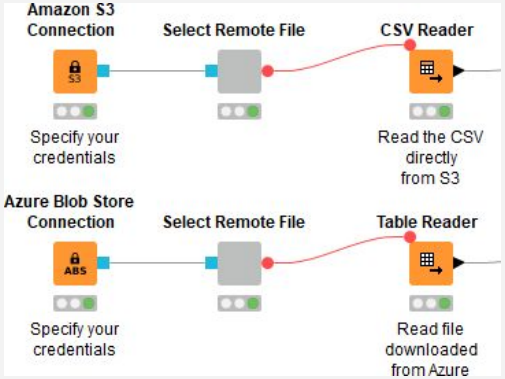


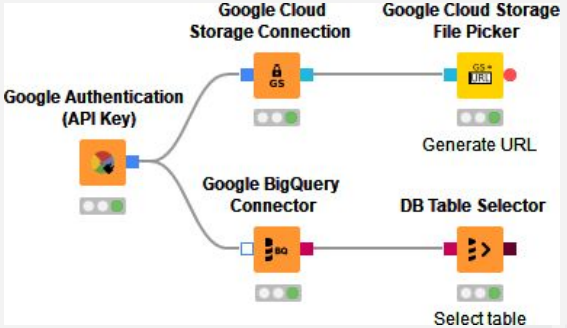


Remote File Handling – Cloud Storage

Integrate remote data sources from Amazon AWS, Microsoft Azure, and Google Cloud

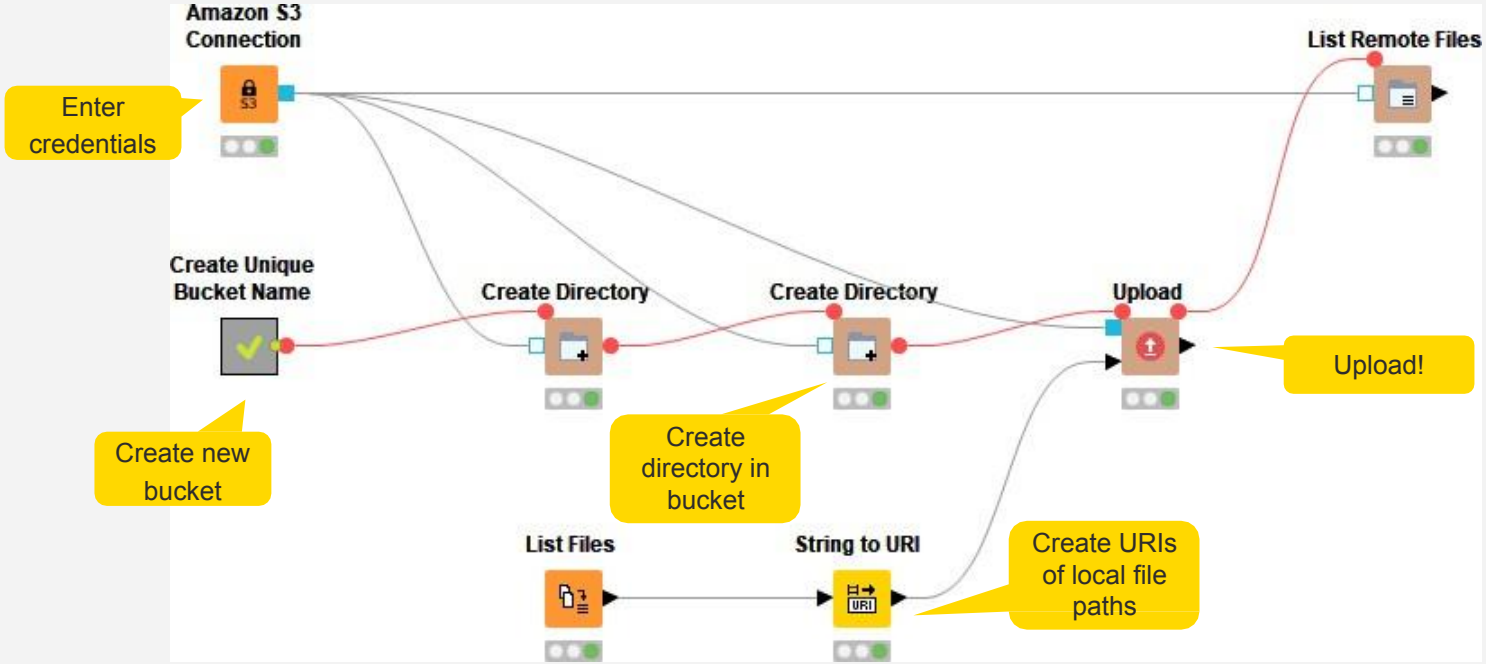
- Upload files
- Download files, or read their content directly into KNIME
- List files in remote directories
- Create directories
- Delete files / directories



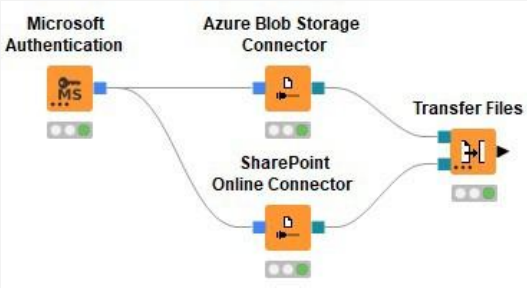


Remote File Handling – Cloud Storage

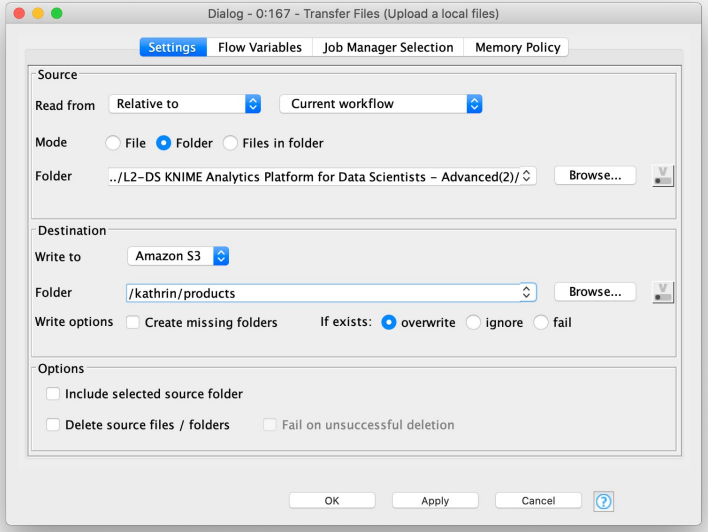
Example: Upload all files from a local directory to Amazon S3

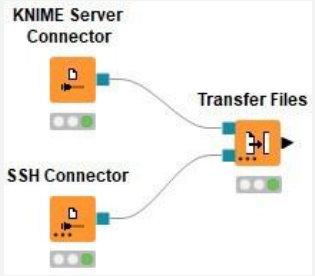


Full Flexibility with the Tranfser Files node

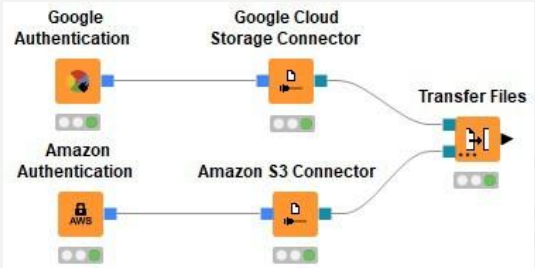


Same cloud environment





On-premise

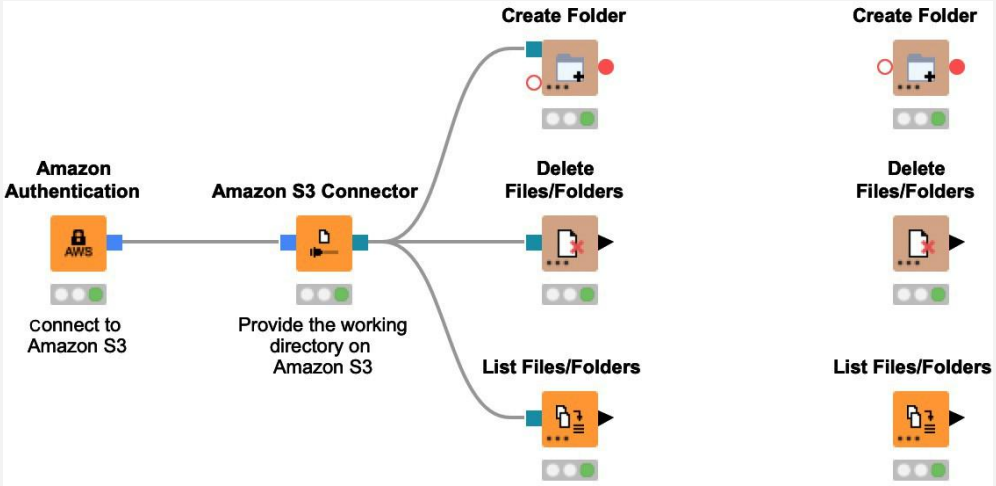


Cross cloud environments

Other Utility Nodes

Can be used local and with remote file systems

- Create a folder
- Delete files or folders
- List all files is a folder



- Further information about file handling https://docs.knime.com/latest/analytics_platform_file_handling_guide/index.html

Considerations when Productionizing a Model

DevelopmentDeployment Maintenance

Reproducibility (data, models)
Training (tuning, autoML)
Testing
Inference time
Model interpretability
Workflow interpretability

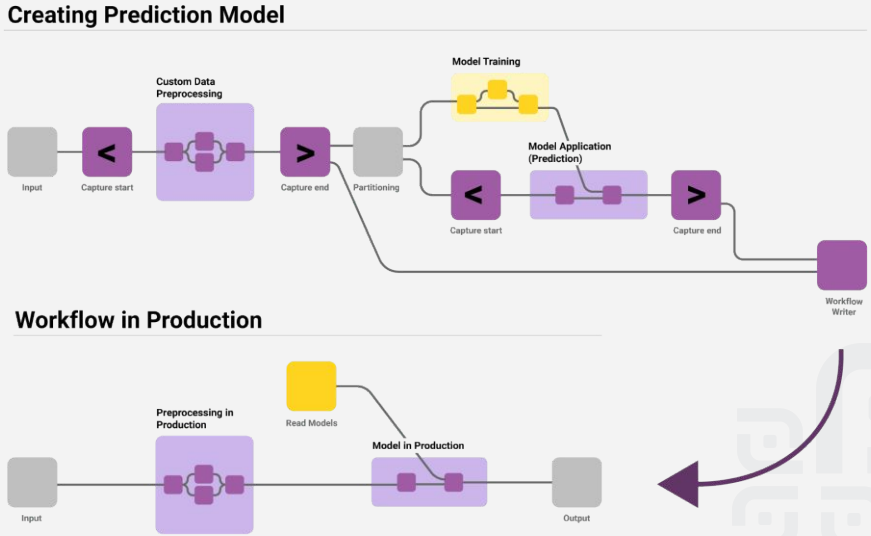
Inference type (batch, online)
Deployment type (data app, REST, schedule, trigger)
Validation (quality assurance)
Testing
Model versioning

Automation

Data and model drifts
Ground truth delay
Continuous availability
Alerts
Retraining
(Re)training time
Testing

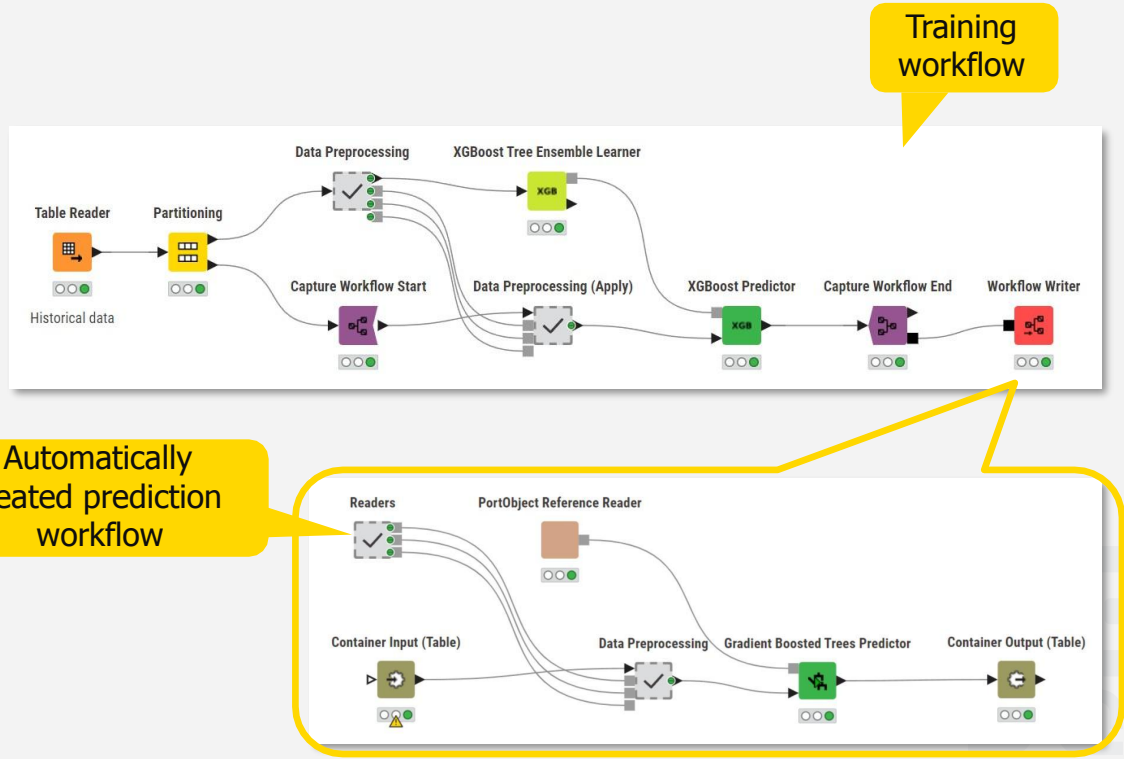
Integrated Deployment

- **Capture core segments of your workflow for reuse**
  - Facilitates deployment of relevant parts of a workflow (e.g., custom data preprocessing, model application for prediction)
- Captured segments are saved automatically as workflows with all the relevant settings and transformations
- Captured segments can be automatically reused with no changes or manual work required



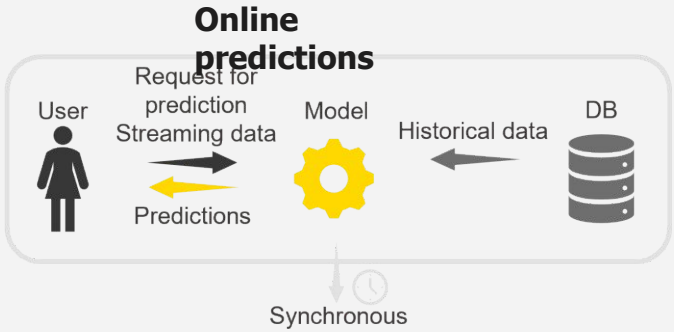
Integrated Deployment

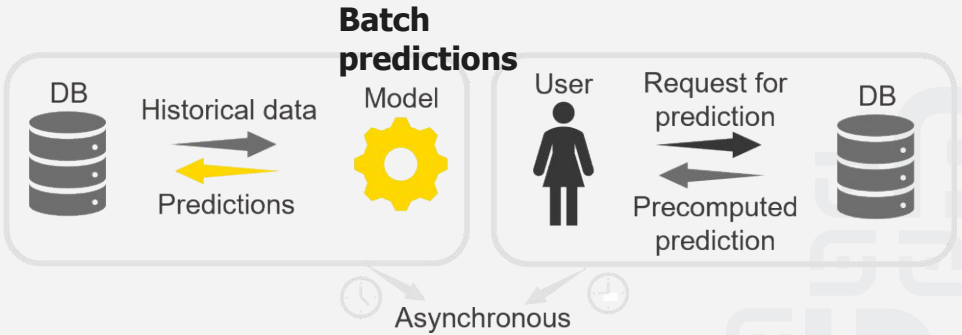
- Create prediction workflows automatically
  - Enable continuous and integrated deployment of a model
- Benefits
  - Time savings
  - Fewer errors
  - Increased compliance
  - Optimized processes
- Other use cases
  - Data pipeline deployment
  - Ad-hoc testing
  - Workflow summary extraction
  - And more



Deployment Option Depends on Prediction Type

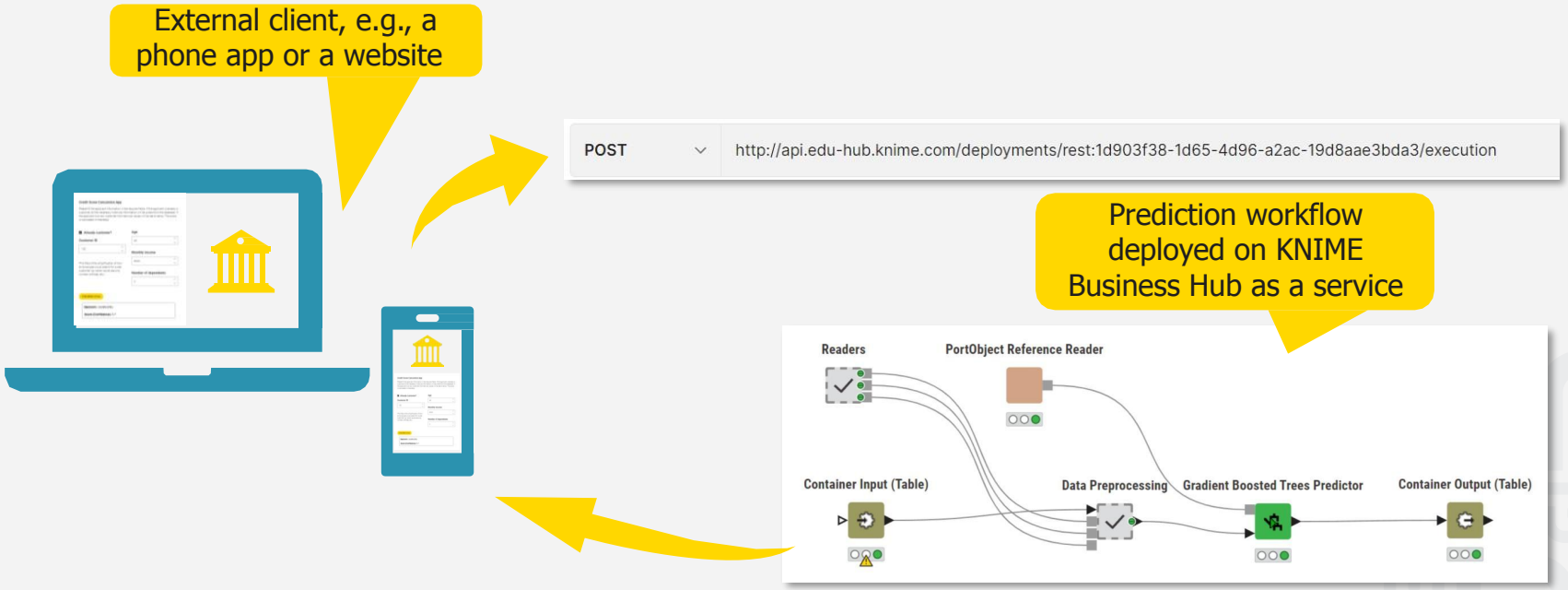
- Online predictions
  - Computed upon a request
  - "Model as a service"
  - With batch features (from historical data) or/and streaming features (from real-time data)
  - **Deploy as a service or as a data app**
- Batch (offline) predictions
  - Precomputed periodically, stored, and are ready to be retrieved upon a request
  - **Schedule or trigger predictions precomputation**





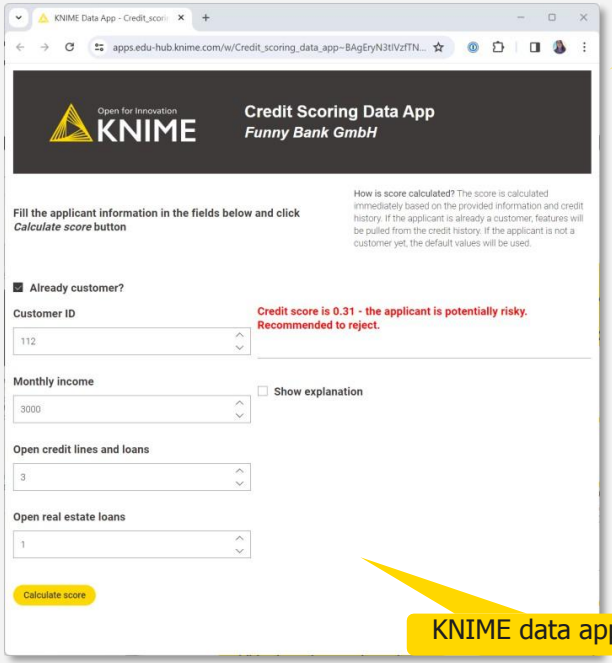
Deploying a Model in a Service

- Allow external clients to interact with a KNIME workflow
  - Use case: a bank allows potential applicants to calculate the credit score **on its website**

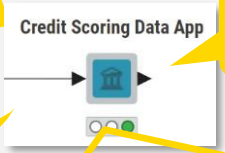


Deploying a Model in a Data App

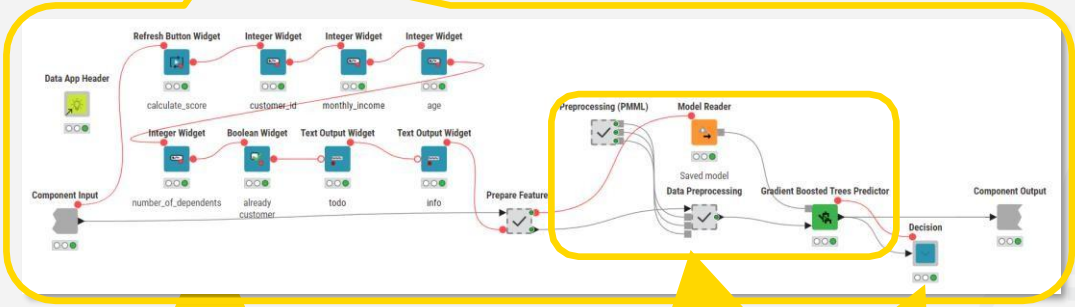
- Enable interaction with a workflow via a user interface in a web browser
  - Use case: a bank employee calculates the credit score for an applicant **via a KNIME data app** shared with an employee



KNIME data app



Underlying workflow deployed on KNIME Business Hub as a data app



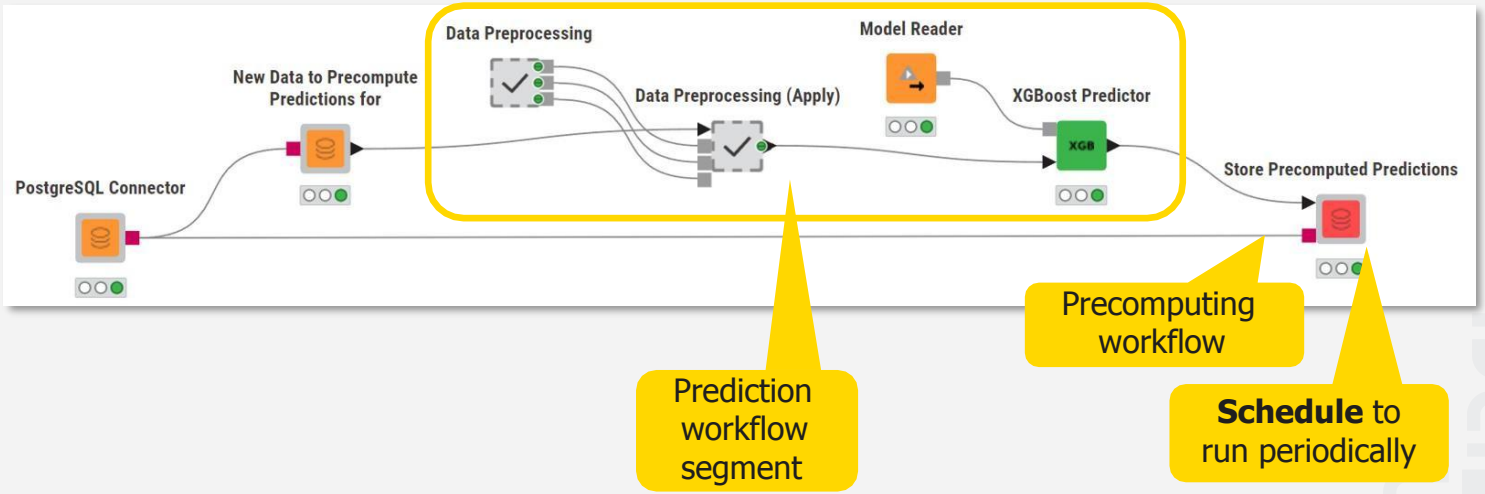
Widget nodes for interaction with the data app

Prediction workflow segment

Visualized predictions

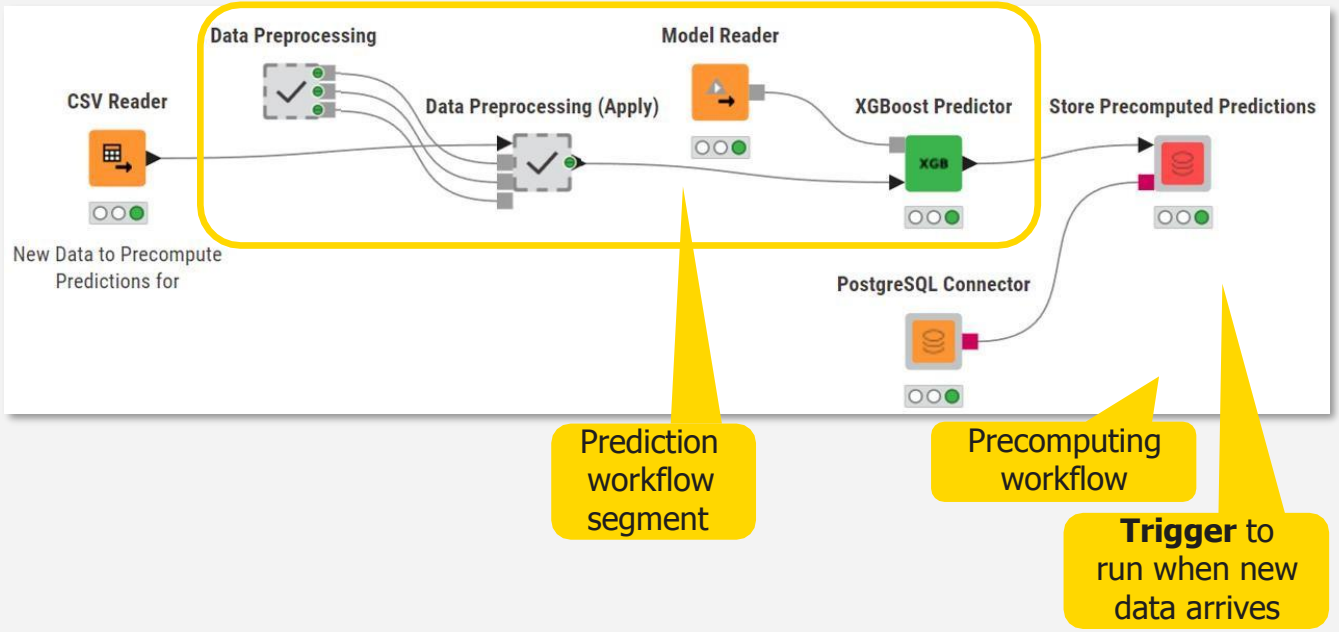
Deploy a Model in a Scheduled Application

- Schedule batch predictions precomputation
  - Use case: a bank **regularly computes** credit scores for all the clients to promote loans to those eligible



Deploy a Model in a Triggered Application

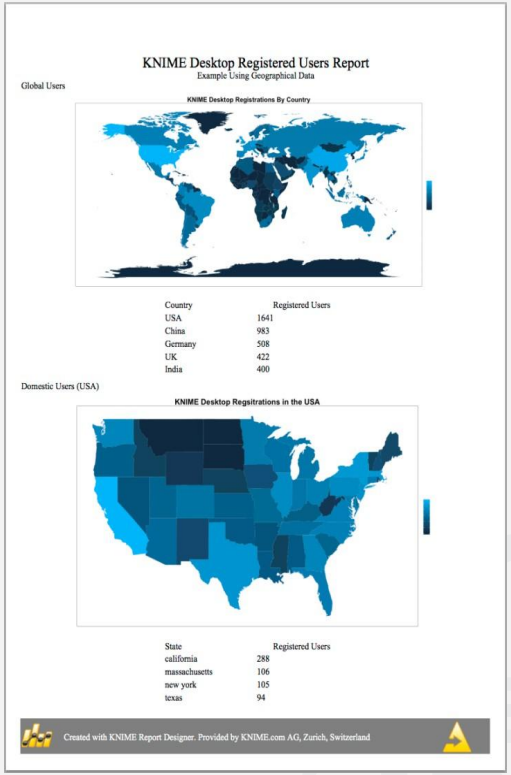
- Trigger batch predictions precomputation
  - Use case: a bank computes credit scores for new clients **once their data arrives** to promote loans to those eligible



Reporting in KNIME

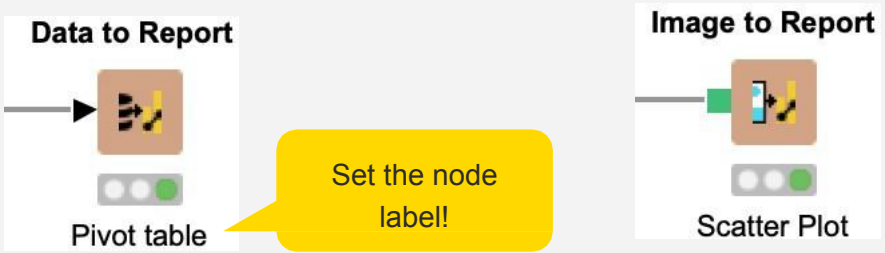
Reporting in KNIME

- Reporting in KNIME is done via a 3rd party application named BIRT (Business Intelligence Reporting Tool)
- Data is sent to BIRT from KNIME using special nodes.
- Reports in BIRT are constructed from report items, which may include images, tables, charts and labels.
- Reports may be generated in a variety of formats (html, pdf, pptx, xlsx, docx, ...)



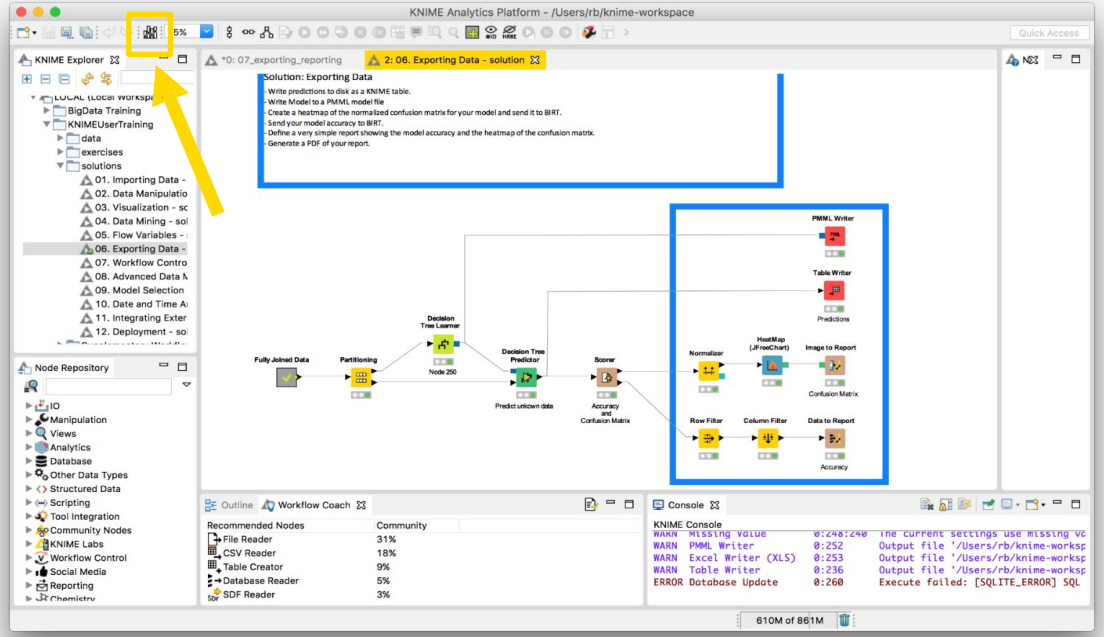
Send Image / Data to Report

- Sends a data table or images to BIRT
- PNG and SVG are supported formats (see node description for details)
- Hint: The node label will be used to identify the data source in the reporting view -> Make sure to use fitting labels if you have more than one data source



Edit the Report

Open the workflow > Click the Report Editor button in the tool bar

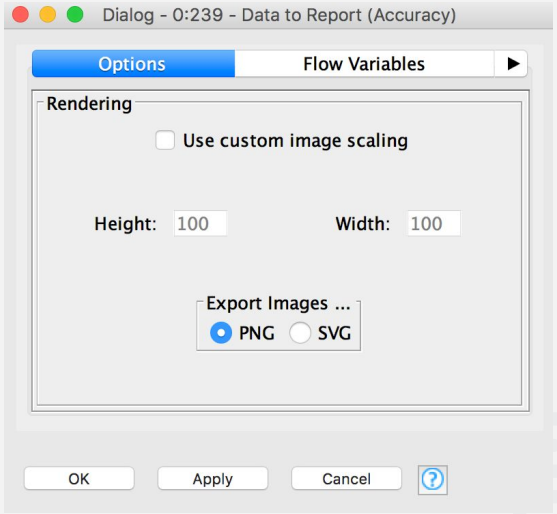


Installation

- Can be installed via KNIME -> Install KNIME Extension
- Install the KNIME Report Designer

New Node: Data to Report

Send a data table to BIRT

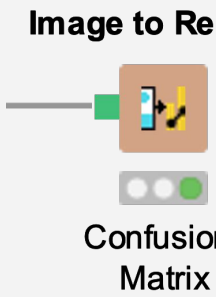


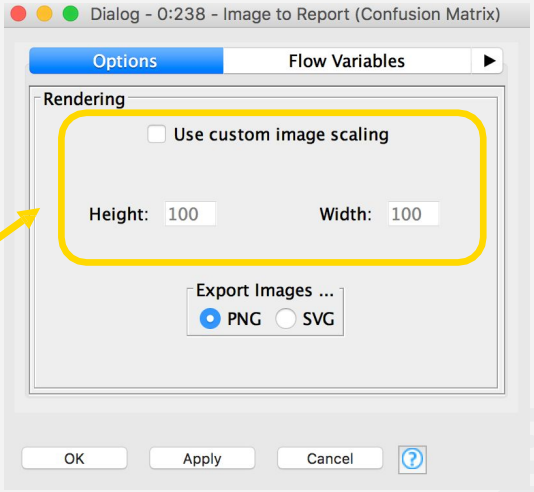
Hint: The node label will be used to identify the data source in the reporting view -> Make sure to use understandable labels if you have more than one data source

New Node: Image to Report

Send an image to BIRT

- PNG and SVG are supported formats (see node description for details)

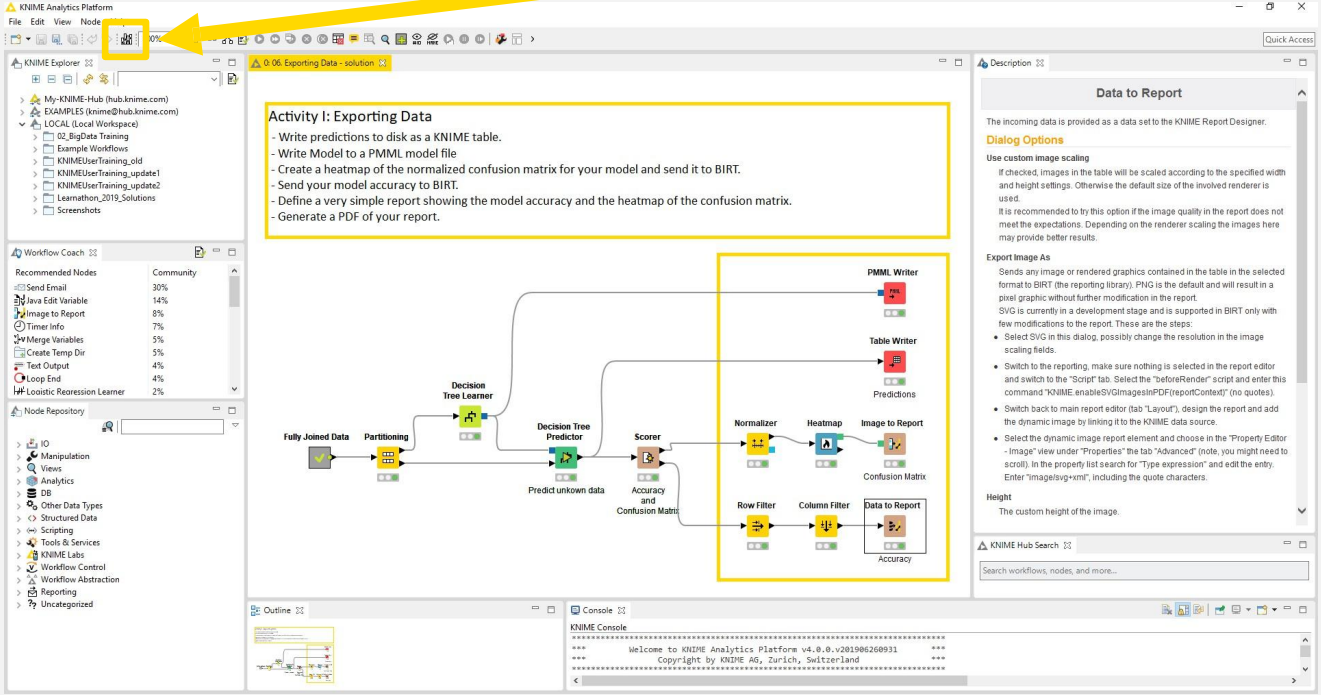




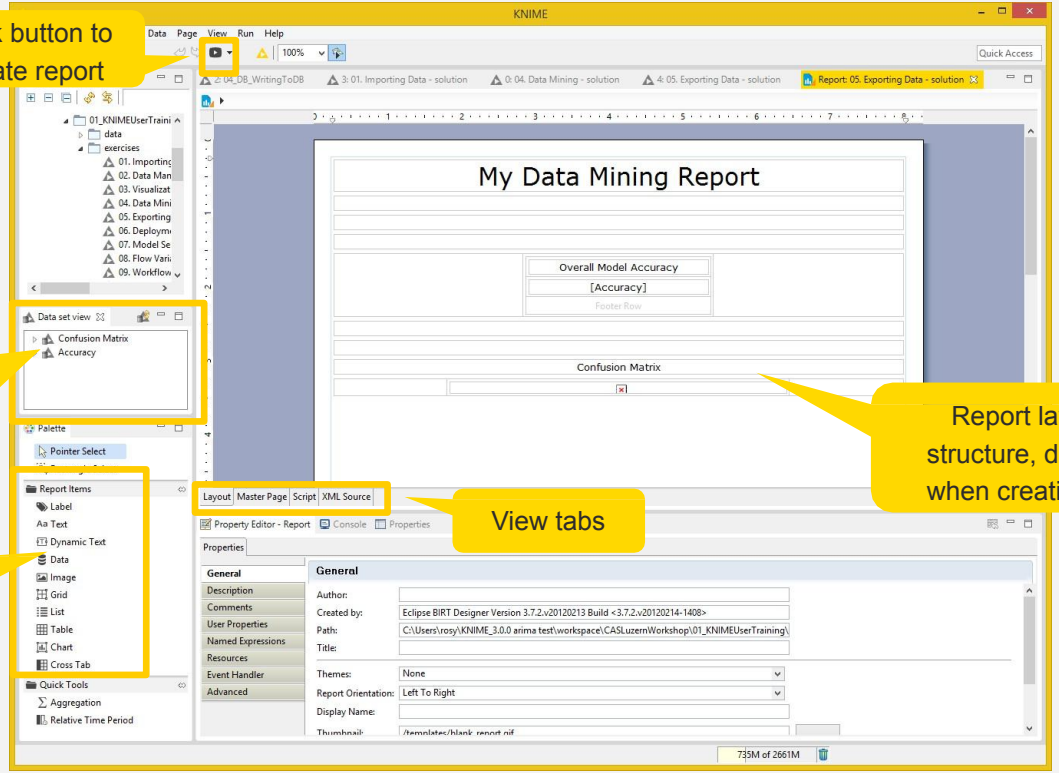
Hint: Customize the image size in the Data to Report node to fit the report

Edit the Report

Open the workflow and click the Report Editor button in the tool bar

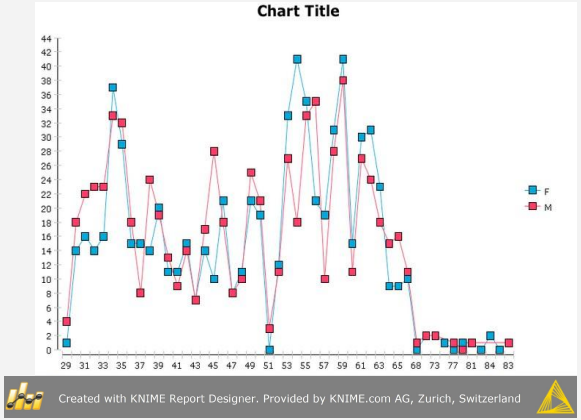


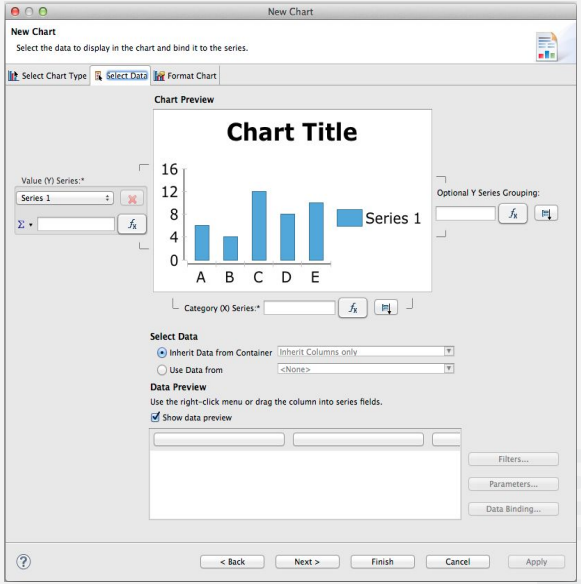
Reporting Perspective



Charting in BIRT

- Many chart types
- Fine control of plot appearance
- Familiar ‘Excel Like’ interface
- Supports interactivity





Tips & Tricks

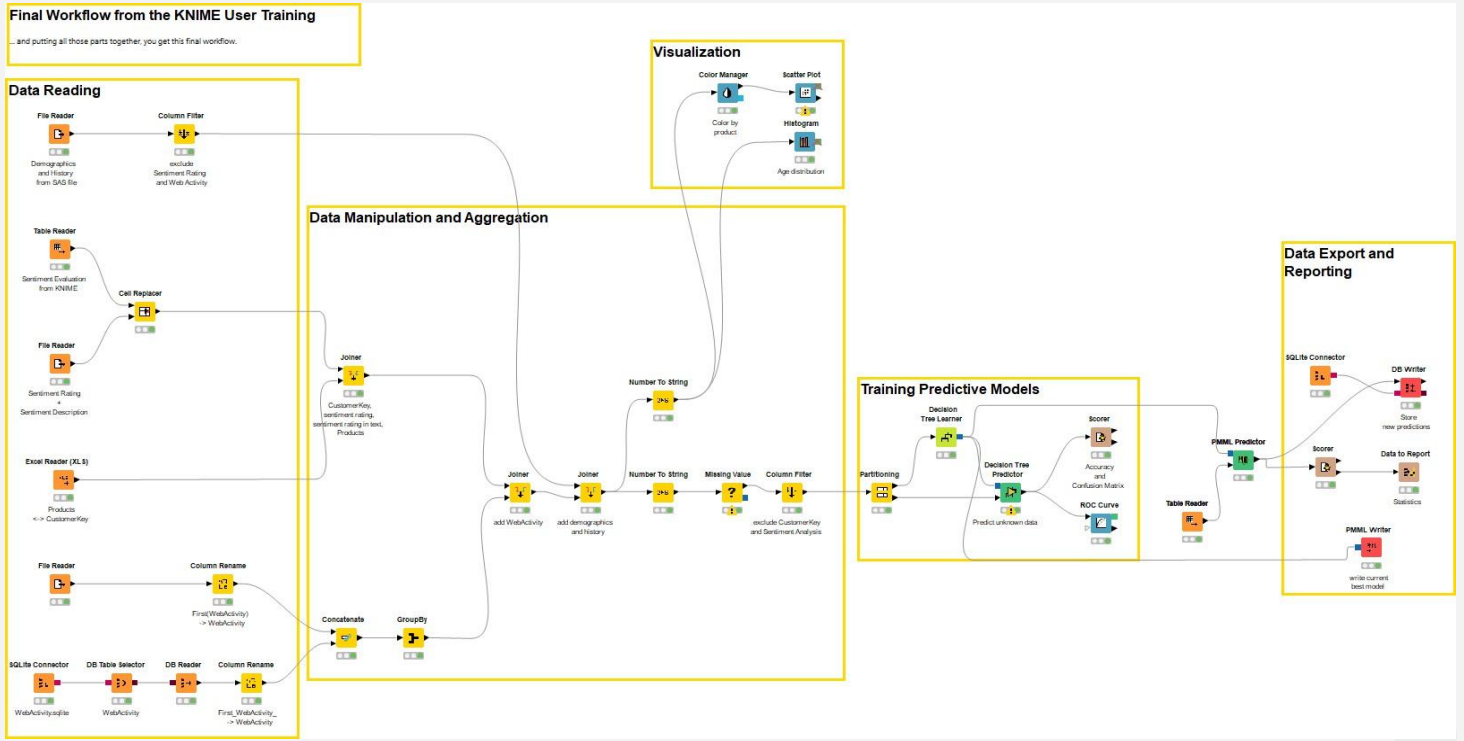
- Use an underlying grid to structure the report
- Names of columns should not change
- Use the grouping function to combine results
- Use the Master Layout Tab (For footers etc.)

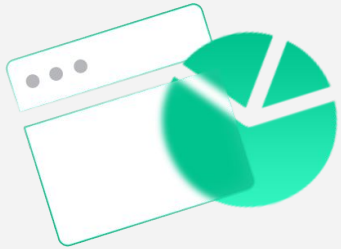
Exporting Data Exercise

Start with exercise: *Exporting Data*

- Write the predictions to a KNIME table
- Write the decision tree model to a PMML model file
- Create a heatmap of the normalized confusion matrix of your model and send it to a BIRT report
- Send your model accuracy to a BIRT report
- Create a simple report showing the overall accuracy and the heatmap of the confusion matrix
- Generate a PDF of your report

Today's Example





Terima Kasih