



Open for Innovation

KNIME

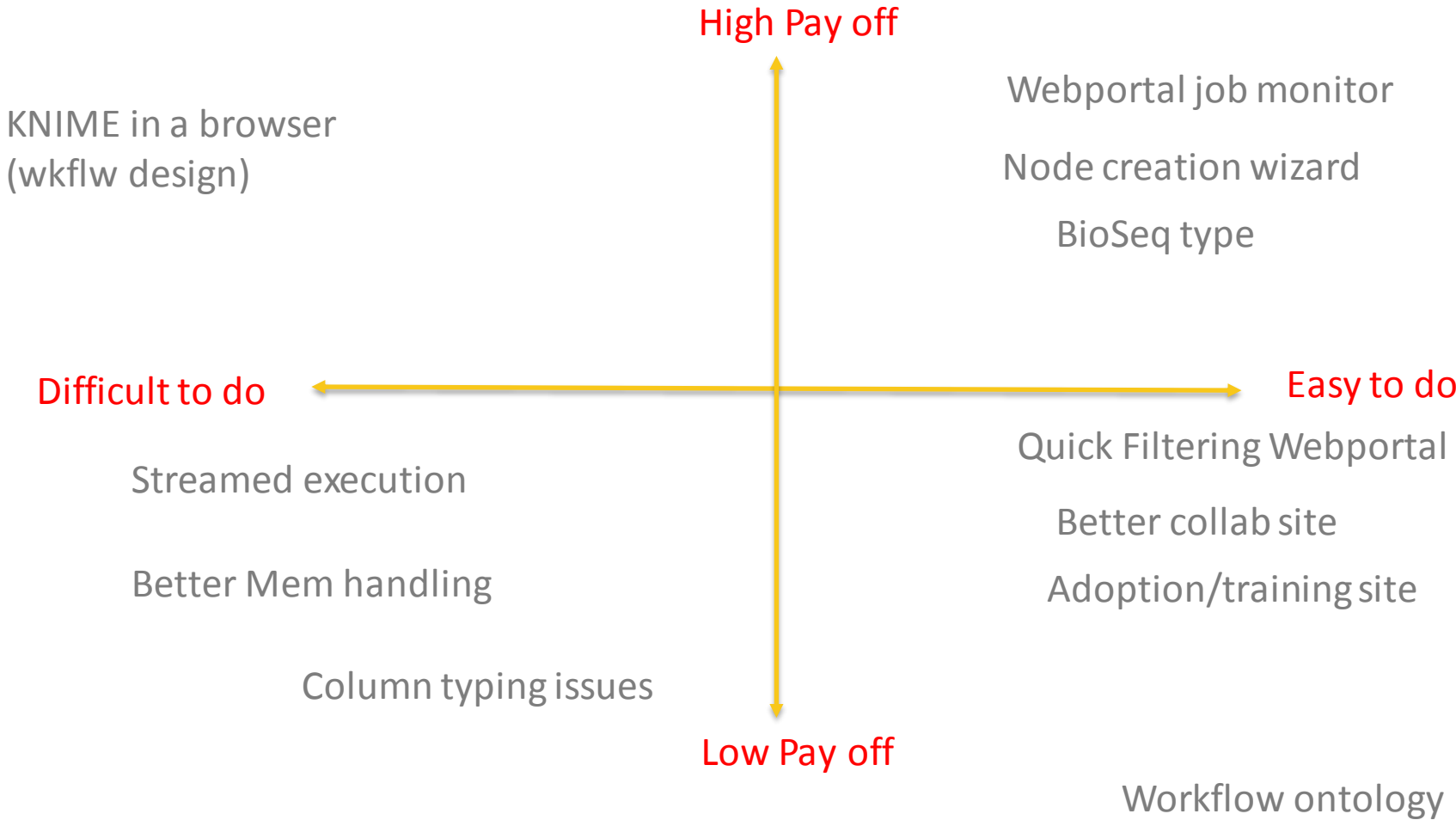
2nd KNIME Chemistry Workshop

Thorsten Meinel

KNIME.com AG, Zurich, Switzerland

Items from last workshop





Items from last workshop I

- Webportal Job Monitor
 - On the list for the new server
- Node creation wizard
 - Need details...
- BioSeq type
 - Easy, need input from users
 - Discussion for partner meeting in November (?)

Items from last workshop II

- Quick filtering Webportal
 - On the list for the new server
- Better memory handling
 - Ongoing process
 - Details of problems would be helpful
- Better collaboration site
 - What about a section in the KNIME Wiki?
- Adoption training site
 - See above

Items from last workshop III

- Column typing issues
 - Should get better with the auto-conversion
- Workflow Ontology
 - Need details

What's cooking



Interactive Python scripting

- Python nodes similar to the Interactive R nodes

The screenshot displays the KNIME Python View interface. At the top, a workflow diagram shows a 'Data Generator' (Node 6) connected to a 'Python Script' (Node 2), which is then connected to a 'Python View' (Node 3). The main window is titled 'Image - 0:3 - Python View' and contains a bar chart with three series: 'Universe_0' (red), 'Universe_1' (cyan), and 'mean' (magenta). The chart shows data for four categories on the x-axis. Below the chart, the Python script is visible, showing the following code:

```
1 import numpy
2 output_table = input_table.copy()
3 output_table = output_table.groupby(
4     'Cluster Membership').median()
5 mean = []
6 for i in range(len(output_table)):
7     mean.append(numpy.mean([
8         output_table['Universe_0_0'][i],
9         output_table['Universe_0_1'][i],
10        output_table['Universe_1_0'][i],
11        output_table['Universe_1_1'][i]]))
12 output_table['mean'] = mean
```

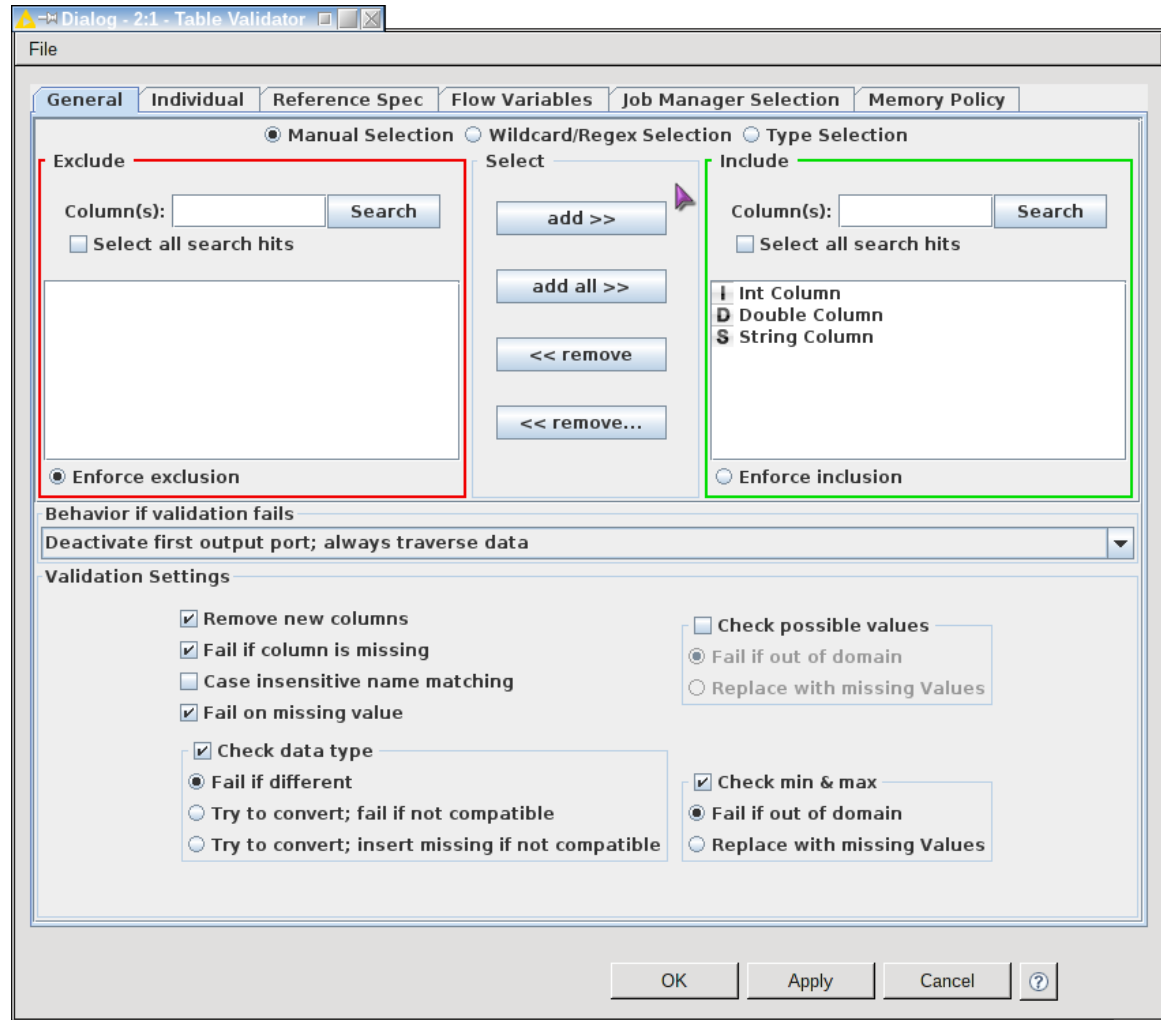
The interface also includes a 'Script' tab with 'Options', 'Flow Variables', and 'Memory Policy' sub-tabs. The 'Flow Variables' tab shows a list of variables: 'i' (int), 'input...' (Data...), 'mean' (list), and 'output...' (Data...). The 'Memory Policy' tab shows a list of memory policies: 'dict', 'int', 'Data...', and 'Data...'. The 'Script' tab shows the following variables: 'Universe_0_0', 'Universe_0_1', 'Universe_1_0', 'Universe_1_1', and 'Cluster Membership'. The 'Flow Variables' tab shows the following variables: 'i', 'input...', 'mean', and 'output...'. The 'Memory Policy' tab shows the following variables: 'dict', 'int', 'Data...', and 'Data...'. The bottom of the interface shows a table with the following data:

Cluster Membership	Universe_0_0	Universe_0_1	Universe_1_0	Universe_1_1	mean
luster_0	0.302621	0.61682	0.530223	0.499106	
luster Membership					0.487192
luster_0					

The bottom of the interface shows a status bar with the text 'Execution successful' and buttons for 'OK', 'Apply', 'Cancel', and a help icon.

Data Validation Node

- Compares incoming table to a reference and reports differences

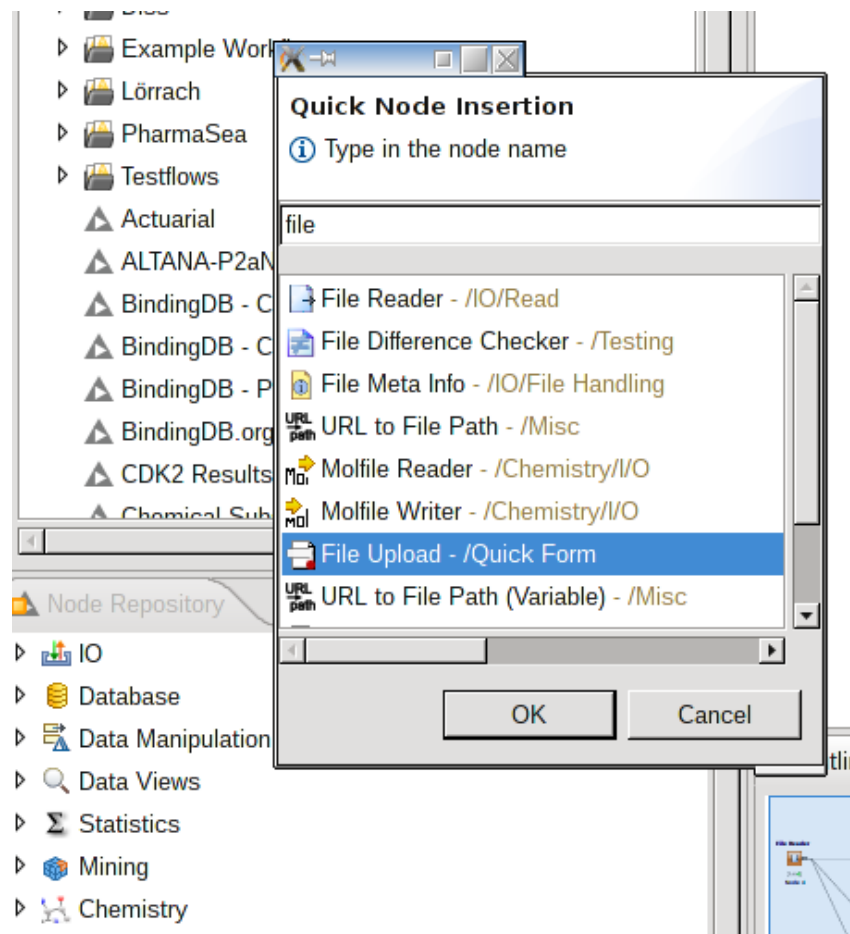


JSON Nodes

- Allow reading, manipulating and writing JSON similar to the XML nodes

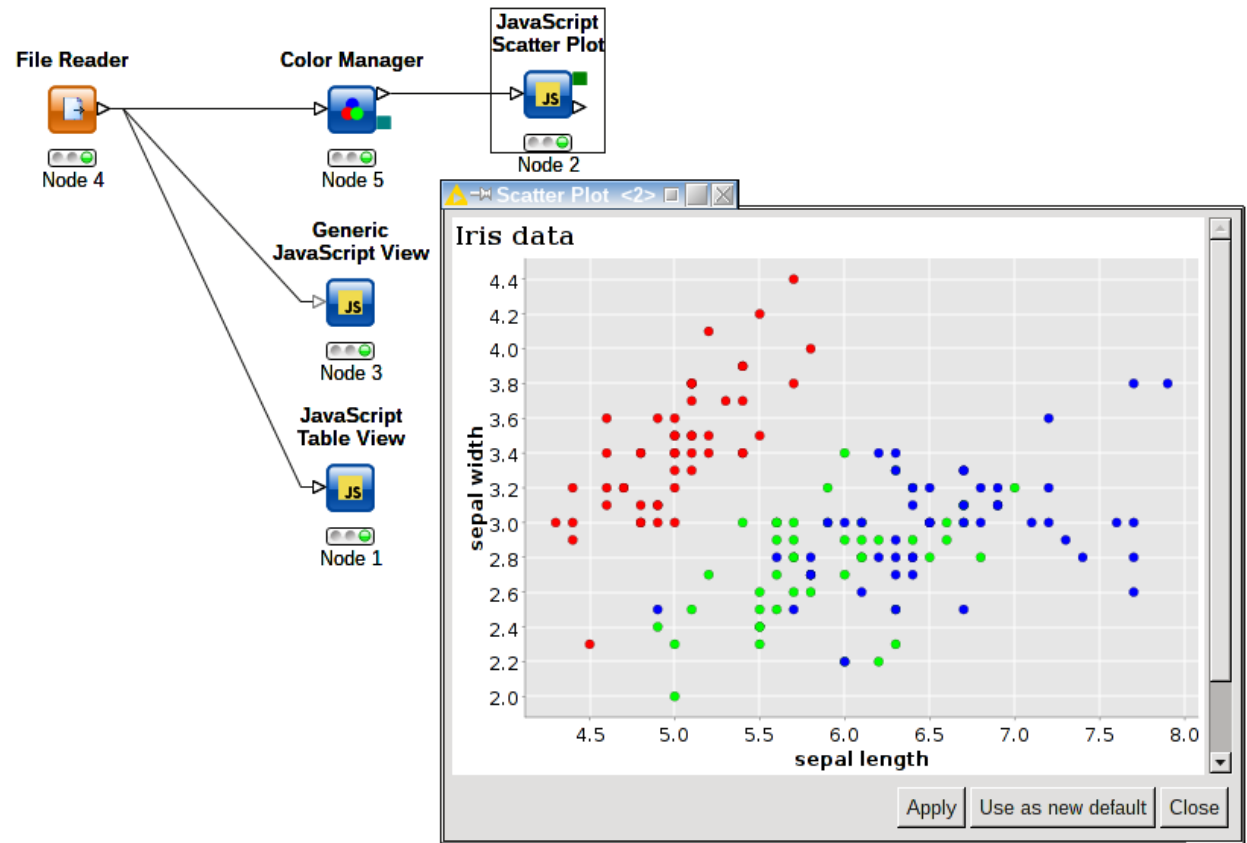
Quick Node Insertion

- Ctrl-Space, Search, Enter



JavaScript Nodes

- Same representation in Client and Webportal
- Interactivity



Server 4.0

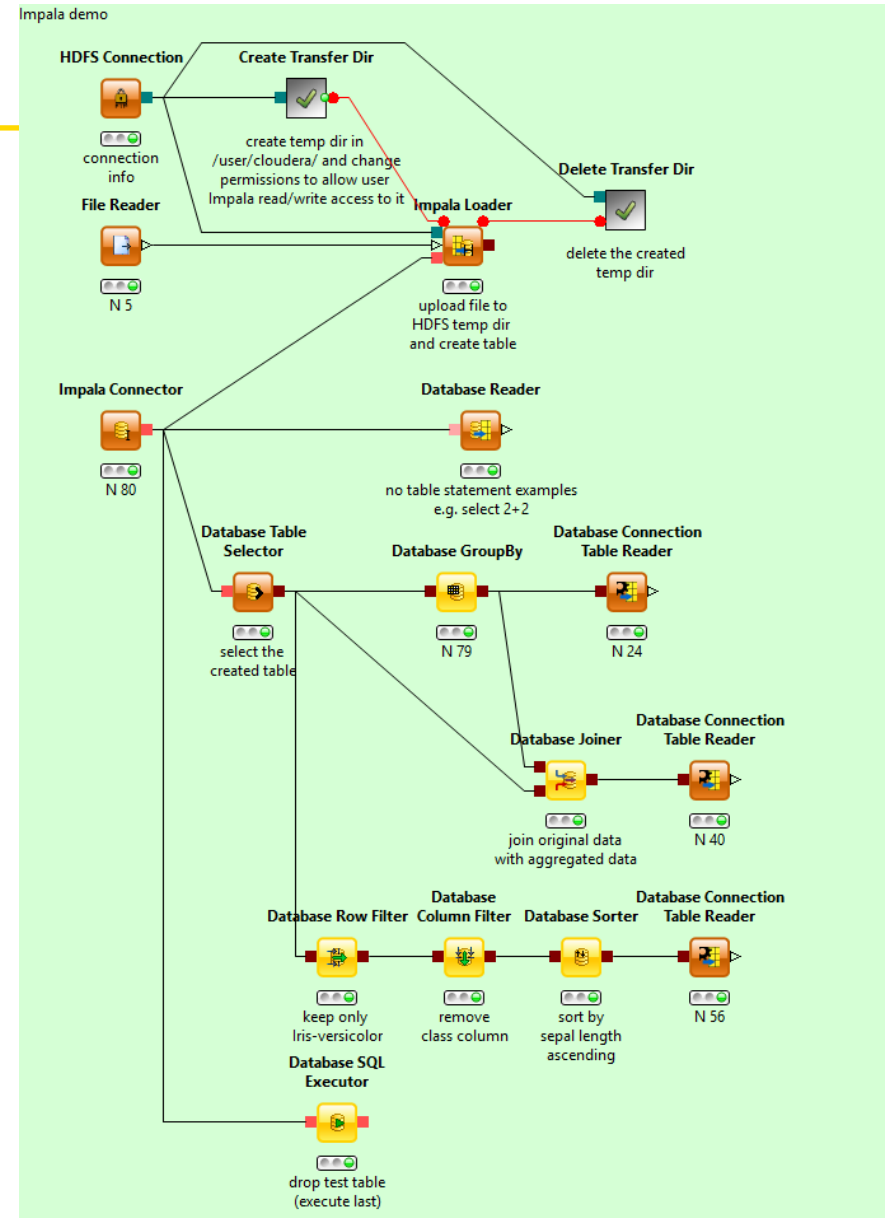
- Based on Tomcat (TomEE)
- REST interface
- All new features added to new server
 - Crontab-like scheduling mechanism
 - Streaming
 - Server-managed client licenses
 - Administrator view
- Old Glassfish-based server will only receive bugfixes

Workflow Diff

- Compare two versions of a workflow
- First step textual
- Second step graphical

BigData Nodes I

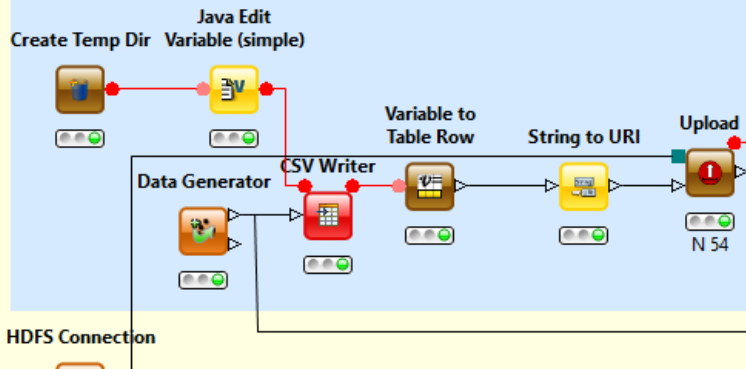
- Connectors to Hive & Impala
 - Use existing DB nodes further down
- Connector to HDFS for file up- and download



BigData Nodes II

HDFS file handling demo

Upload csv file to HDFS file system



HDFS Connection



Download csv file from HDFS file system

