



3rd KNIME SIG ChemInf - Update from KNIME -

Thorsten Mehl
KNIME.com AG, Zurich, Switzerland

Items from last workshop / Wishlist



Items from last workshop/Wishlist I

- Easier node creation for developers
 - Currently working on declarative JS nodes
- Still open:
 - Workflow preferences
 - Fulltext search in node repository
 - Integration of GLMol in WebPortal
 - More than one report per workflow
 - Reports distributable in meta/subnodes
 - Easier molecule input quickform
 - "Linked" workflows in local workspace that can be synchronized with server
 - Select all nodes between two selected nodes in workflow editor

What's cooking

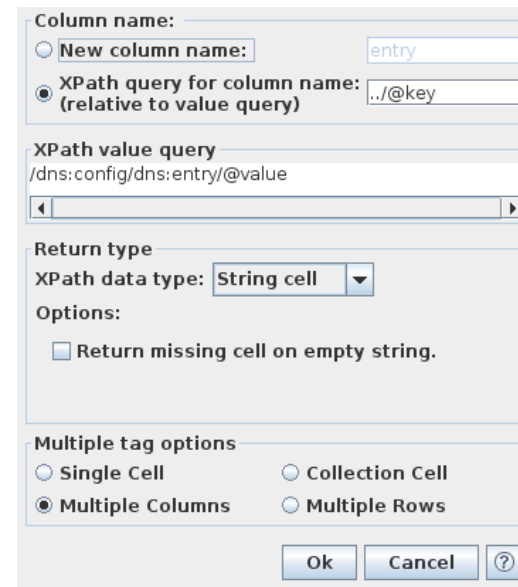
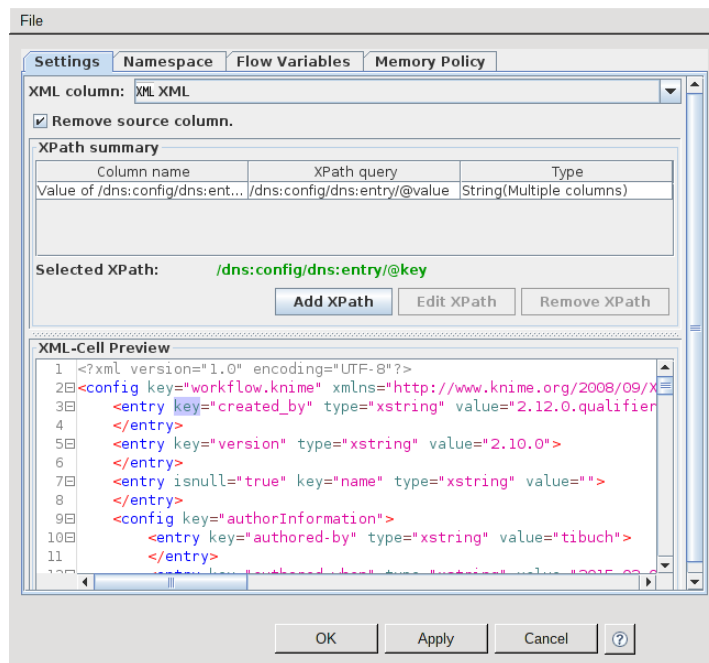


Java 8 & Eclipse 4.4

- Update to Java 8 Update 45 (very likely)
- Update to Eclipse 4.4
- Update to BIRT 4.4
 - No inline report preview any more
- Ideally the users don't notice anything (yet)
- Developers can use Java 8 features
- Preview available at
<https://tech.knime.org/forum/knime-general/knime-212-technology-preview-feedback-requested>

New XPath Node

- Allows multiple XPath expressions
- Column names computed from XPath expressions
- Paths can be generated from XML preview



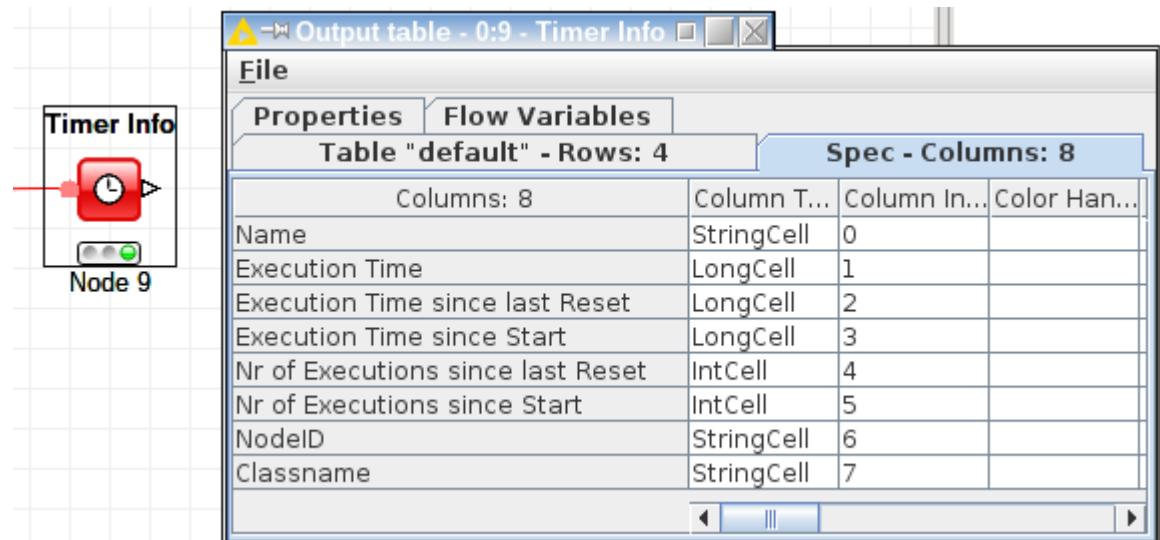
New JSON Nodes

- Columns to JSON
 - Creates a JSON object for each row with the column values as properties
- JSON Path (multi)
 - Multiple JSON Path expressions per node
- JSON to Table
 - Extract table structures from a JSON object

- ...and others
 -  Columns to JSON
 -  JSON Column Combiner
 -  JSON Combine and Write
 -  JSON Path (multi)
 -  JSON Row Combiner
 -  JSON to Table
 -  JSONPath (Dictionary)

Node Timer

- Creates table with execution times of all nodes in the workflow








The image shows a screenshot of the KNIME software interface. On the left, a 'Timer Info' node (Node 9) is visible, represented by a red square with a clock icon. A red arrow points to the node. On the right, a window titled 'Output table - 0:9 - Timer Info' is open, displaying a table with 8 columns and 4 rows. The table has tabs for 'Properties' and 'Flow Variables'. The table content is as follows:

Table "default" - Rows: 4			
Columns: 8	Column T...	Column In...	Color Han...
Name	StringCell	0	
Execution Time	LongCell	1	
Execution Time since last Reset	LongCell	2	
Execution Time since Start	LongCell	3	
Nr of Executions since last Reset	IntCell	4	
Nr of Executions since Start	IntCell	5	
NodeID	StringCell	6	
Classname	StringCell	7	

- Global Node Timer
 - Collects information for the whole KNIME instance

Other new nodes

- Sleep/Pause
 - Shamelessly “stolen” from Vernalis
- Node Timer
 - Outputs execution statistice for each node in the workflow
- Python with DB ports
- MongoDB
 - Read, Write, Update JSON

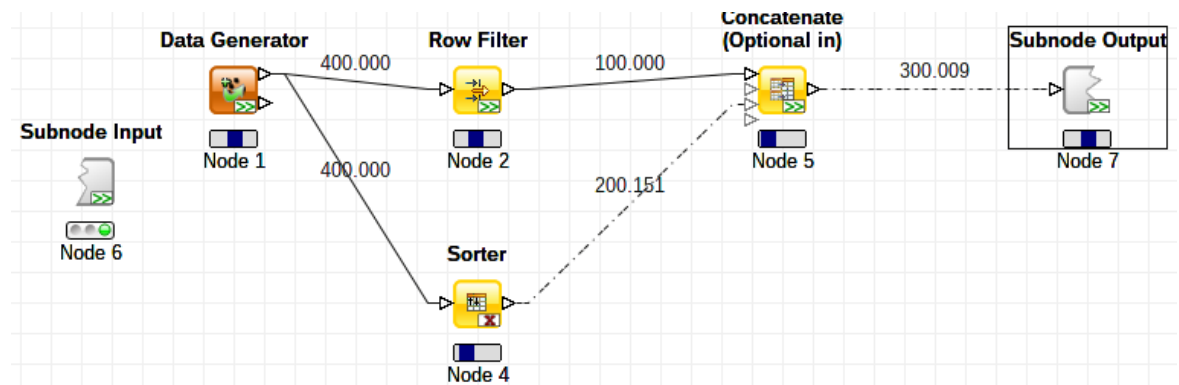
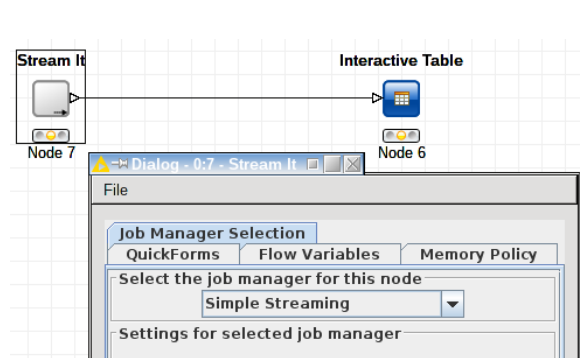
-  MongoDB Reader
-  MongoDB Remove
-  MongoDB Save
-  MongoDB Update
-  MongoDB Writer

Javascript Views & Nodes

- Frame for automatically creating nodes from a JavaScript file and a node description

Streaming Executor

- Labs project, not ready for production!
- Data rows are streamed to subsequent nodes as soon as they become available
- Not all nodes as streamable
- Output data not visible after execution



REST interface for KNIME Server

- List repository contents
- Workflow execution via REST calls
 - Load workflow
 - Set parameters
 - Execute workflow (multiple times)
 - Get results and/or report

WebPortal Templates

- Layout can be configured by templates
 - Footer & header
 - Main panel
 - Login page
 - ...
- Custom stylesheet
- Custom JavaScript libraries
 - Can be re-used in JS-based views

Spark/MLLib integration

- Based on Spark MLLib
- Scalable machine learning library
- Runs on Hadoop
- Algorithms for
 - Classification (decision tree, naïve bayes, ...)
 - Regression (logistic regression, linear regression, ...)
 - Clustering (k-means)
 - Collaborative filtering (ALS)
 - Dimensionality reduction (SVD, PCA)

