


Data Access and Database





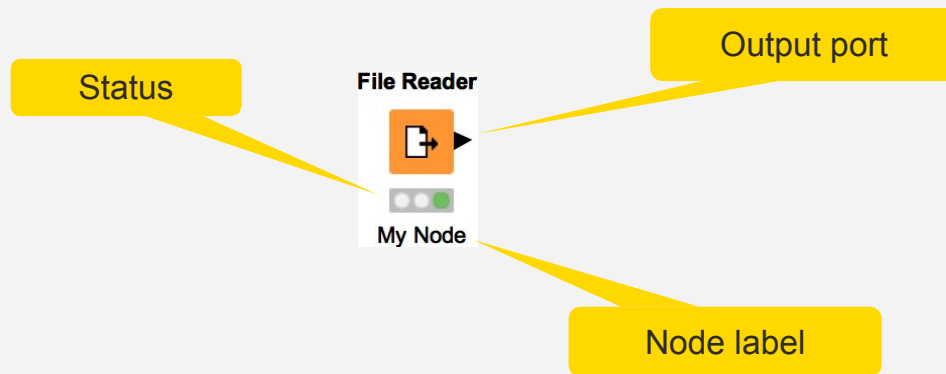
Importing Data Accessing Files



Data Source Nodes

Typically characterized by:

- Orange color
- No input ports, 1-2 output ports



New Node: File Reader

Workhorse of the KNIME Source nodes

- Reads all text based files (e.g. csv, txt, etc.)
- Many advanced features allow it to read most 'weird' files
 - Short lines, inline comments, headers and special encoding



YouTube KNIME TV Channel video:

<https://youtu.be/flaHQw-Qhlg>

File Reader Configuration

The screenshot shows the 'File Reader (My Node)' configuration dialog. It has tabs for 'Settings', 'Flow Variables', 'Job Manager Selection', and 'Memory Policy'. The 'Settings' tab is active, showing options to enter an ASCII data file location, basic settings like 'read row IDs' and 'read column headers', and a preview of the data. Callouts point to the 'File path' input field, the 'Basic Settings' section, the 'Advanced...' button, the 'Preview' table, and the 'Help Button' at the bottom right.

File path

Basic Settings

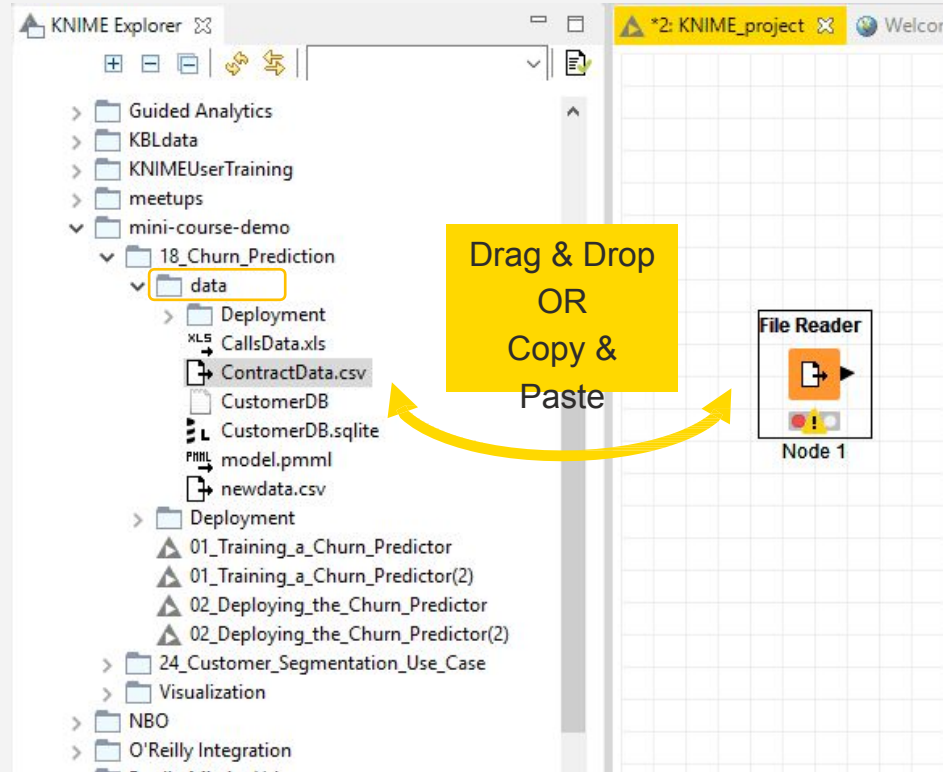
Advanced Settings

Preview

Help Button

Row ID	D Sepal...	D Sepal...	D Petal...	D Petal...	S Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3	1.4	0.1	setosa
14	4.3	3	1.1	0.1	setosa
15	5.8	4	1.2	0.2	setosa
16	5.7	4.4	1.5	0.4	setosa
17	5.4	3.9	1.3	0.4	setosa
18	5.1	3.5	1.4	0.3	setosa

Alternative Faster Way ...



Filenames and the knime:// Protocol

■ Absolute URL

Input location

`knime://LOCAL/KNIMEUserTraining/data/Sentiment%20Analysis.table` Browse...

■ Mountpoint-relative URL

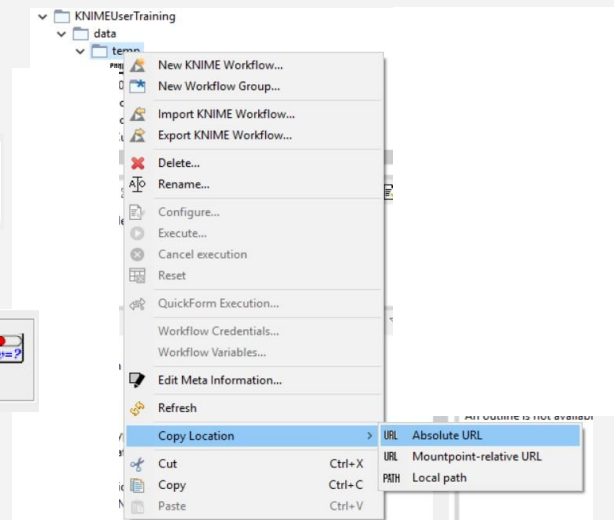
Input location

`knime://knime.mountpoint/KNIMEUserTraining/data/Sentiment%20Analysis.table` Browse...

■ Local path

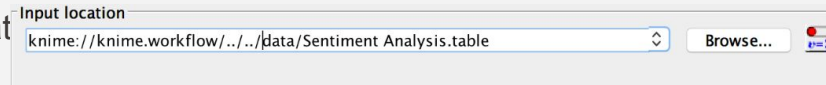
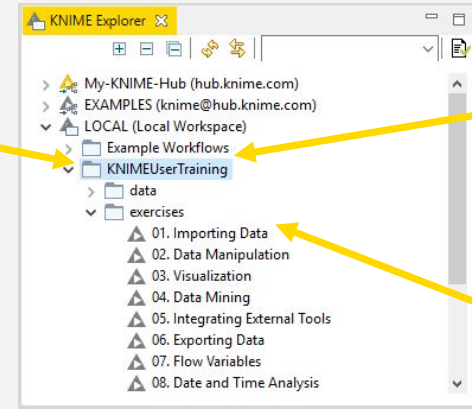
Input location

`/Users/rb/knime-workspace/KNIMEUserTraining/data/Sentiment Analysis.table` Browse...



Workflow-Relative File Paths

- Best choice if workflows are to be shared
- Requires matching folder structure within workflow group
 - Independent of environment outside of workflow group
- Example: Path to „Sentiment Analysis.table“
 - Local path:
C:\Users\rb\knime-workspace\KNIMEUserTraining\data\Sentiment Analysis.table
 - Workflow relat



YouTube KNIME TV Channel:
<https://youtu.be/U9sP4g4yGwY>

New Node: Simple File Reader

- Faster compared to the File Reader node
- Only basic settings

Simple File Reader



Basic settings

Preview

Dialog - 0:1210 - Simple File Reader

Settings | Advanced Settings | Limit Rows | Encoding | Flow Variables | Memory Policy

Input location
cher/knime-workspace-dw-course/L1-DW KNIME Analytics Platform for Data Wranglers - Ba ⌵ Browse...

Connection timeout [s] 1 ⌵

Reader options

Format

Autodetect format ⚙

Column delimiter \n Row delimiter

Quote char Quote escape char

Comment char

☒ Has column header ☐ Has row ID

☐ Support short data rows

Preview

The suggested column types are based on the first 50 rows only. See 'Advanced Settings' tab.

Row ID	Accou...	Churn	Int'l Plan	VMail ...	State	Area ...	Phone
Row0	128	0	0	1	KS	415	382-4657
Row1	107	0	0	1	OH	415	371-7191
Row2	137	0	0	0	NJ	415	358-1921
Row3	84	0	1	0	OH	408	375-9999
Row4	75	0	1	0	OK	415	330-6626
Row5	118	0	1	0	AL	510	391-8027
Row6	121	0	0	1	MA	510	355-9993
Row7	147	0	1	0	MO	415	329-9001
Row8	117	0	0	0	LA	408	335-4719
Row9	141	0	1	1	WV	415	330-8173
Row10	65	1	0	0	IN	415	329-6603
Row11	74	0	0	0	RI	415	344-9403
Row12	168	0	0	0	IA	408	363-1107
Row13	95	0	0	0	MT	510	394-8006
Row14	63	0	0	0	IA	415	366-0338

OK Apply Cancel ?

File path

New Node: Excel Reader (XLS)

- Reads .xls and .xlsx file from Microsoft Excel
- Supports reading from multiple sheets

Excel Reader (XLS)



Read Excel Sheet Names (XLS)



File system

Sheet specific settings

Preview

Filenames and the knime:// Protocol

■ Local File System

Select file to read:

☒ File ☐ Files in folder

Read from: Local File System

File: 'Users/kathrinmelcher/knime-workspace-data-science-course/Example Workflows/TheData/Customers/CallsData.xls' Browse...

■ Relative to ...

Read from: Relative to

- Current mountpoint
- Current workflow data area
- Current workflow

File: Calls_data.xlsx Browse...

■ Mountpoint

Read from: Mountpoint LOCAL

File: /Example Workflows/TheData/Customers/CallsData.xls Browse...

■ Custom URL

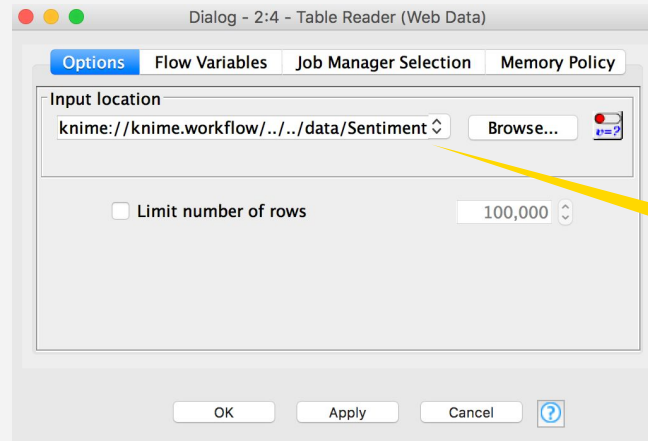
Read from: Custom URL

URL: knime://knime.workflow/data/Calls_data.xlsx Browse...

New Node: Table Reader

- Reads tables from the native KNIME Format.
- Maximum performance, minimum configuration

Table Reader

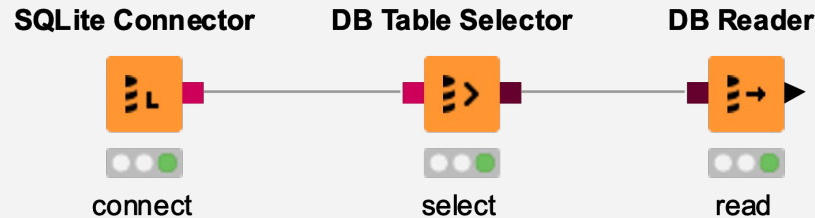


YouTube KNIME TV channel video:

<https://youtu.be/tid1qi2HAOo>

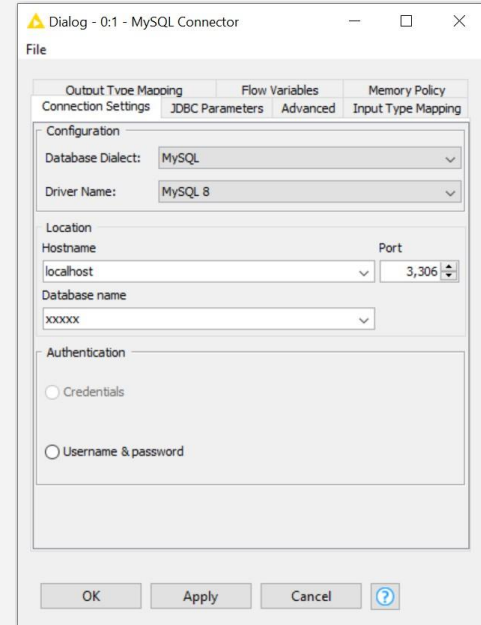
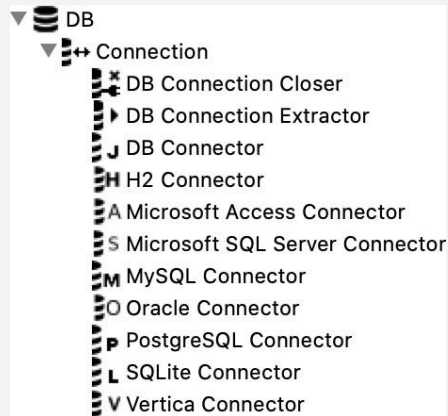
Database Connectivity

- Read data from any JDBC enabled database
- Write your own SQL or model it using dedicated nodes



New Nodes: Database Connectors

- Native: Postgres, MySQL, MS SQL Server, SQLite
- DB Connector (e.g. DB2, HANA).
- Big Data: HIVE and Impala



Importing Data Exercise

Start with exercise: *Importing Data*

Read the following files

- *Sentiment Analysis.table*
- *Sentiment Rating.csv*
- *Product Data2.xls*

Optional: Read the *web_activity* table from the database
WebActivity.sqlite

(hint: drag and drop the files from the KNIME Explorer panel to get started)

You can download the training workflows from the KNIME Hub:

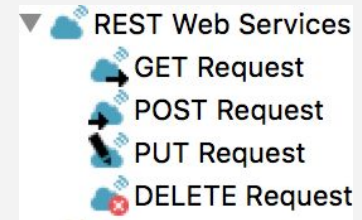
<https://hub.knime.com/knime/spaces/Education/latest/Courses>

/



RESTful Web Services

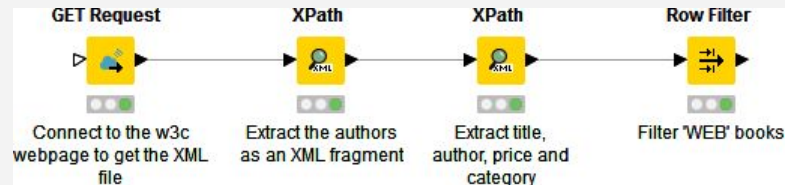
- Use KNIME nodes to interact with RESTful web services
- Send requests using standard HTTP methods



JSON Response:



XML Response:



RESTful Web Services

Enter URL, or
use from column

Add delay between
individual requests

Provide authentication
if necessary

GET Request



Dialog - 2:540 - GET Request

Connection Settings Authentication Request Headers Response Headers

☒ URL:

☐ URL column:

☐ Delay (ms):

Concurrency:

SSL

☐ Ignore hostname mismatches

☐ Trust all certificates

☐ Fail on connection problems (e.g. timeout, certificate errors, ...)

☐ Fail on http errors (e.g. page not found)

☒ Follow redirects

Timeout (s)

Body column:

OK - Execute Apply Cancel ?

[https://www.knime.com/blog/a-restful-way-to-find-and-retrieve-dat](https://www.knime.com/blog/a-restful-way-to-find-and-retrieve-data)

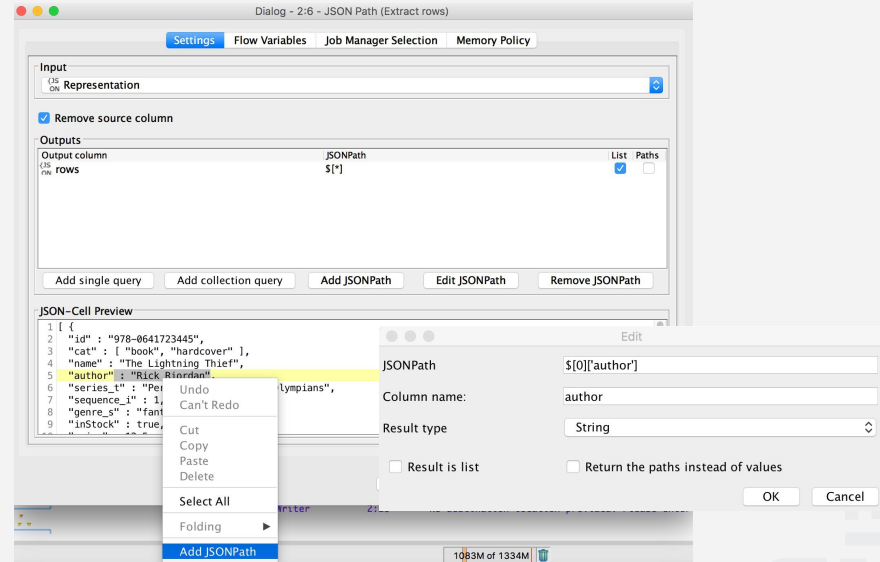
[a](https://www.knime.com/blog/a-restful-way-to-find-and-retrieve-data)

<https://www.knime.com/blog/OSM-meets-CSV-file-and-Google-AP>

!

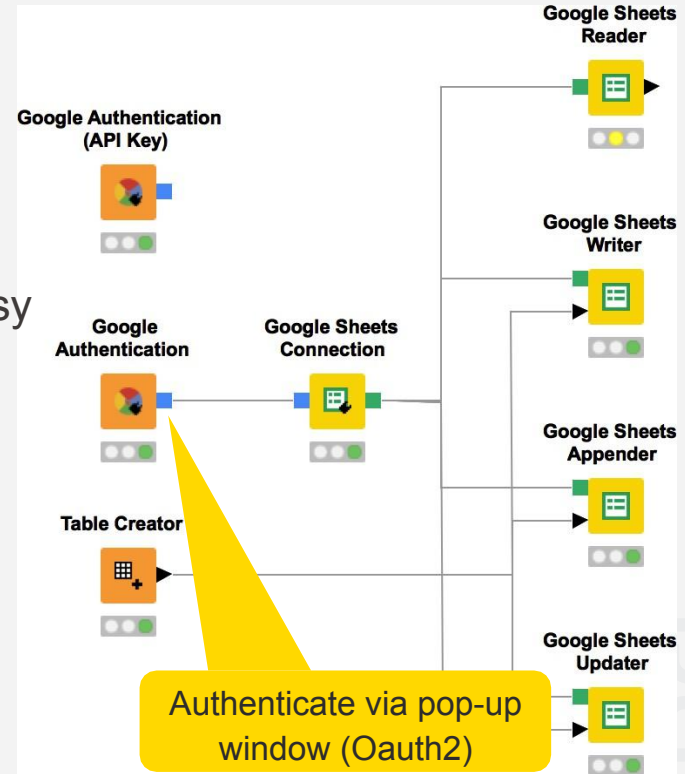
JSON Reader and JSON Path nodes

- Use the JSON Reader (or GET Request) node to get a JSON cell
- Use the JSON Path node to query the JSON file and extract parameters
- Editor window simplifies construction of JSON queries by auto-generating them (click on properties)



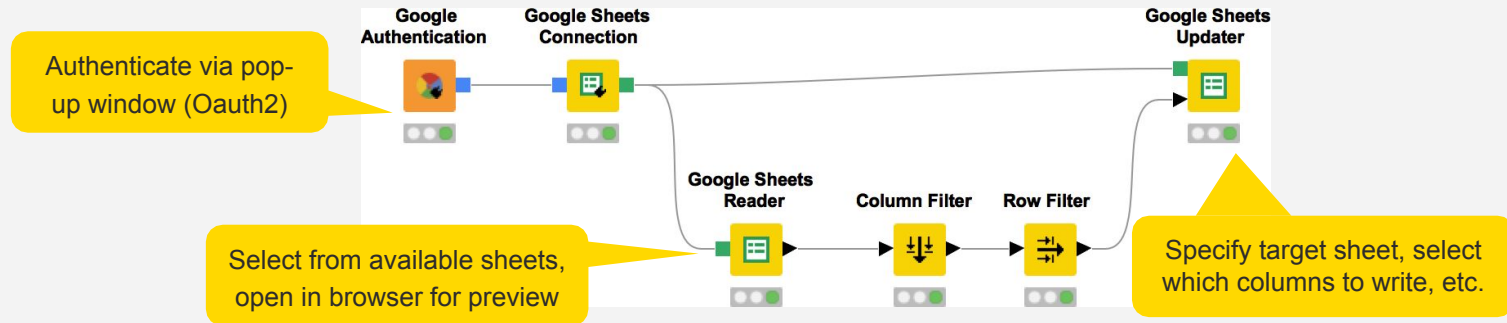
Google Sheets

- Access your data stored in Google Services
 - Read data from Google Sheets
 - Write data to new sheets
 - Modify existing sheets
- Makes collaboration and sharing of data easy
 - (especially vs. sending Excel sheets via email...)



Google Sheets

- Select from available sheets on Google Drive
- Transform data in KNIME, or enrich with new data
- Create new sheet or update existing sheets
 - Allows to read from / write to specific range of sheet (e.g. A1:G10)

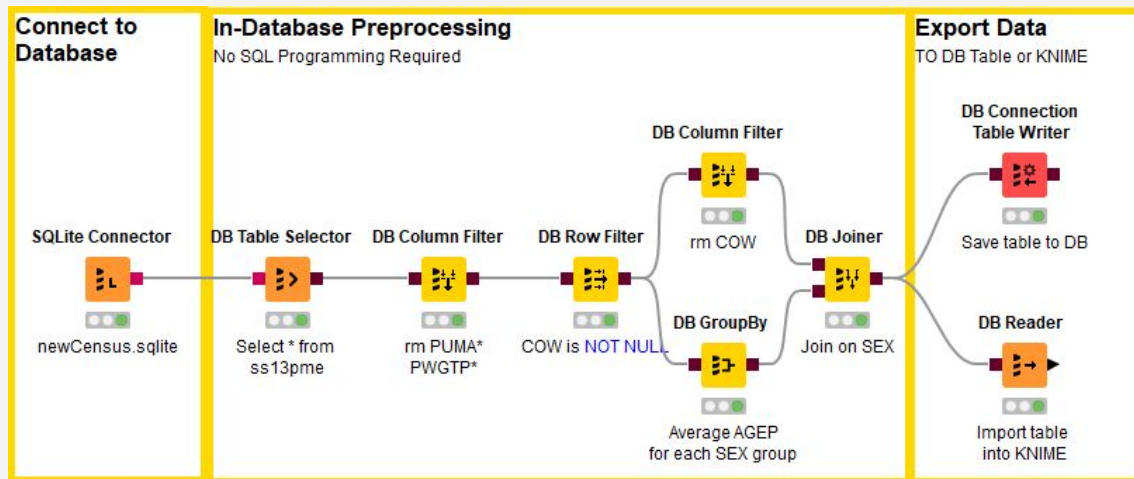


Databases

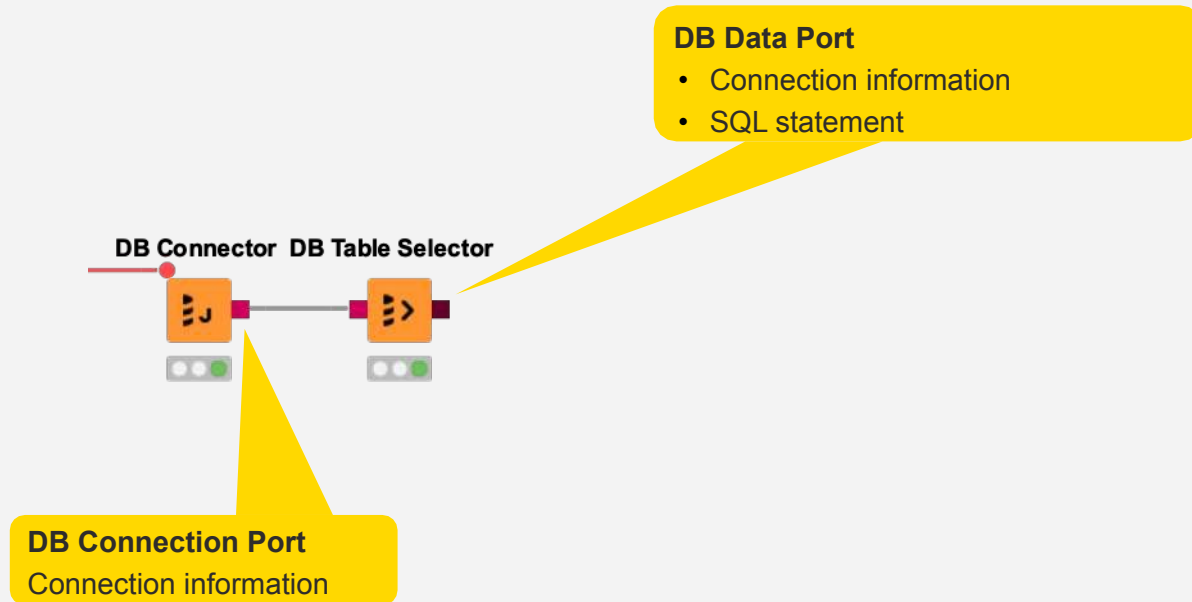
Database Extension

usually assemble complex SQL statements (no SQL coding needed)

- Connect to all JDBC-compliant databases
- Harness the power of your database within KNIME
- Complete rewrite in KNIME Analytics Platform 4.0

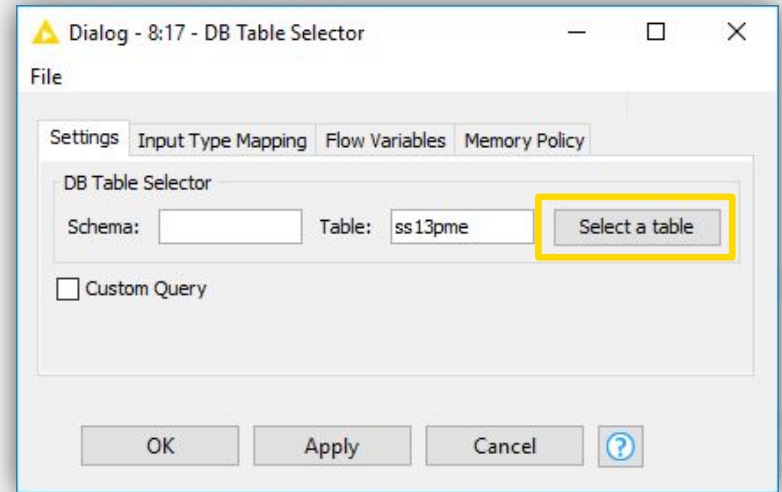
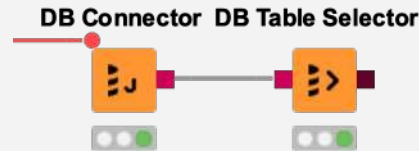
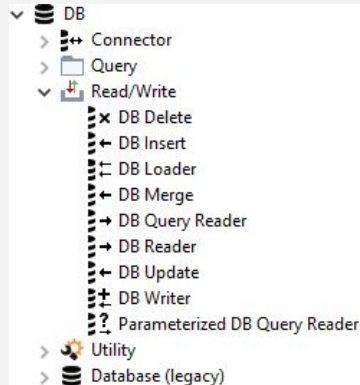


Database Port Types



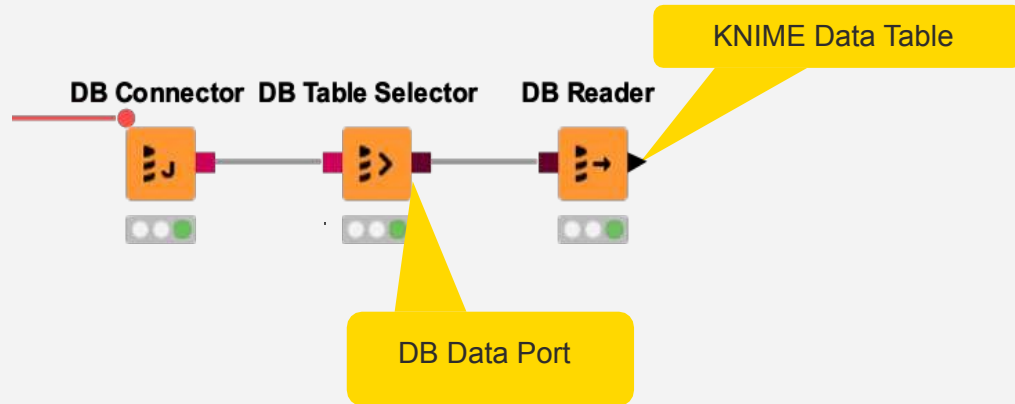
DB Table Selector

- Takes connection information and constructs a query
- Explore DB metadata
- Outputs a SQL query



DB Reader

- Executes incoming SQL Query on Database
- Reads results into a KNIME data table



Database Connectors

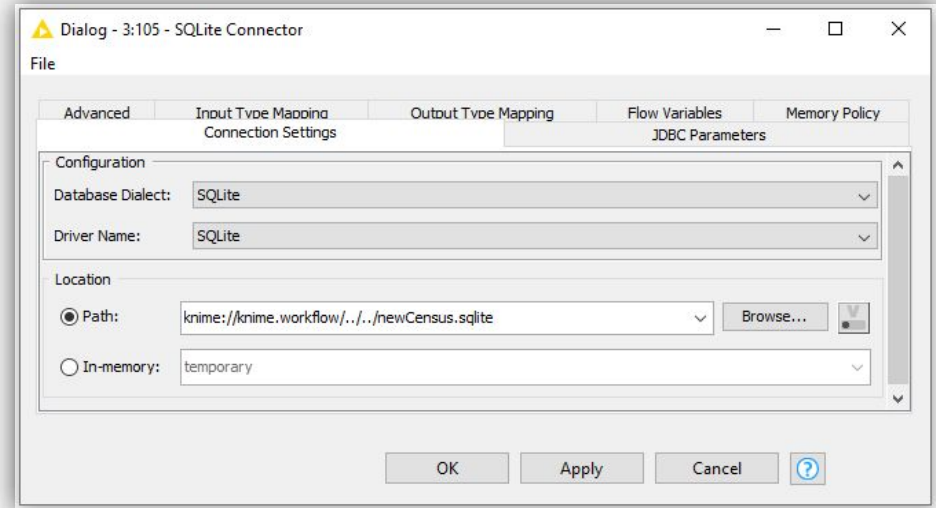
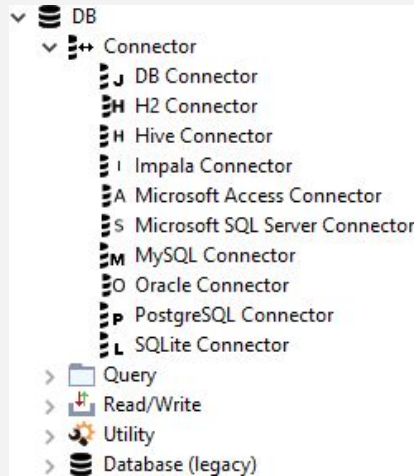
- Dedicated nodes to connect to specific Databases
 - Necessary JDBC driver included
 - Easy to use
 - Import DB specific behavior/capability
- Hive, Impala connectors part of the KNIME Big Data Connectors extension
- General DB Connector
 - Can connect to any JDBC source
 - Register new JDBC driver via
File -> Preferences -> KNIME -> Databases



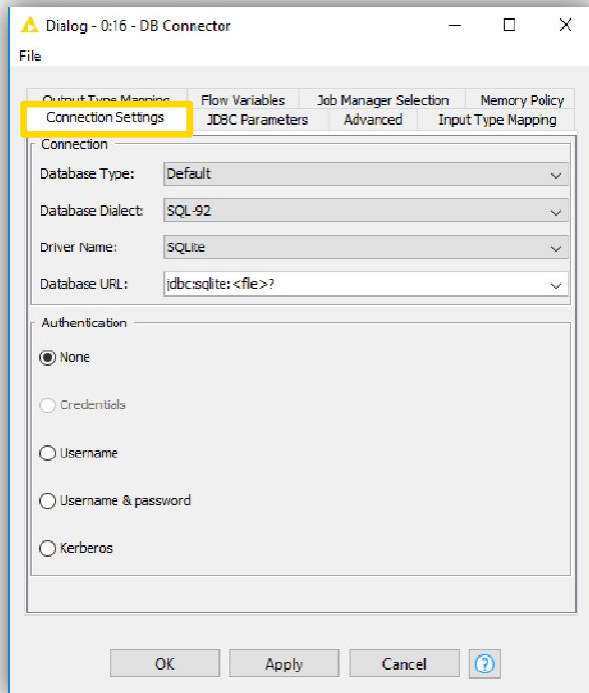
Dedicated Database Connectors

- MySQL, MS SQL Server, Postgres, SQLite, Amazon Redshift, etc.
- Propagate connection information to other DB nodes

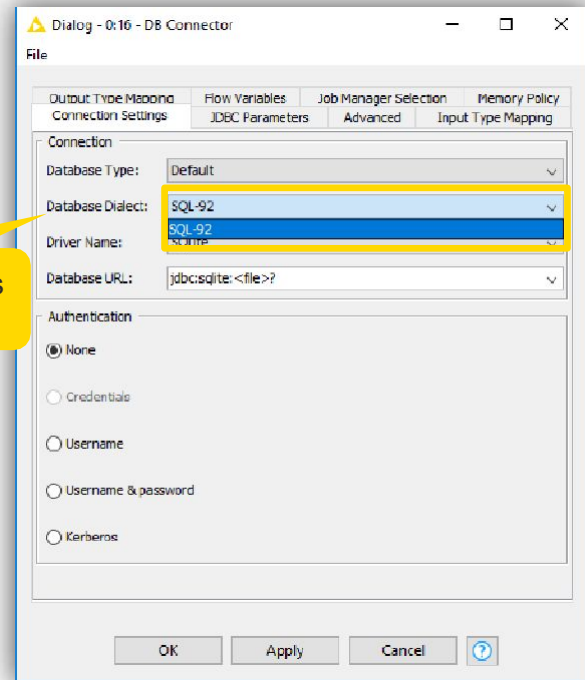
SQLite Connector



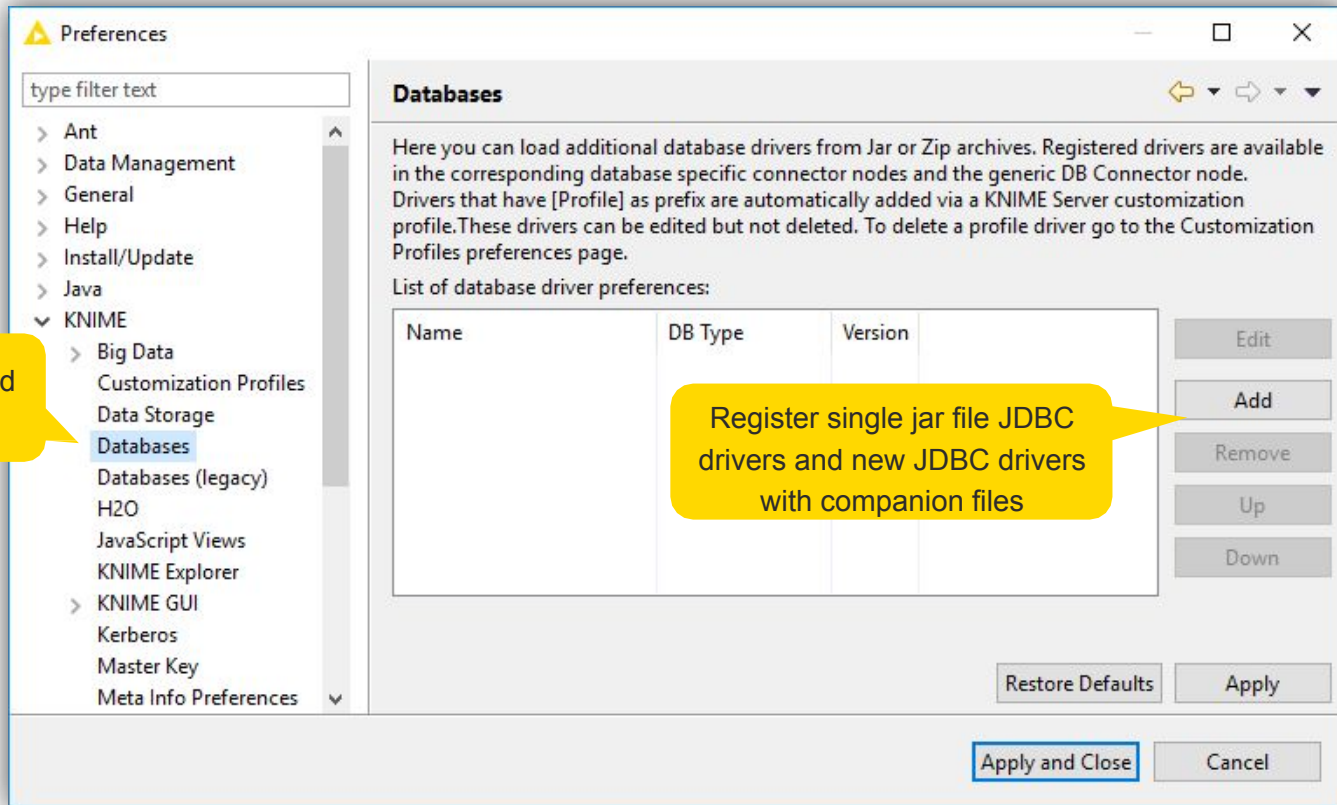
«General» DB Connector Node



Database type defines SQL dialect



Register JDBC Driver



Other Useful Data Sources

■ KNIME Analytics Platform provides many more options for data:

- PMML Reader – reads standard predictive models
- XML Reader with XPATH support
- Python/R Source nodes
- Tika Parser – extracts textual data from 200+ file types
- REST Web Service



- Find out more in by downloading the free book “Will they blend”

<https://www.knime.com/knimepress/download-will-they-blend>

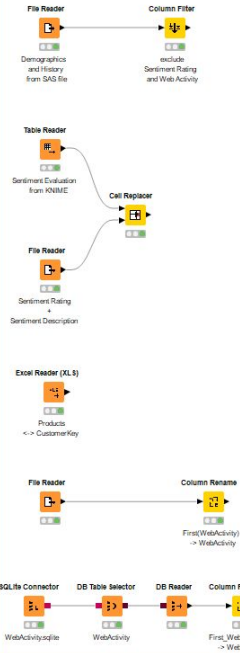


Today's Example

Final Workflow from the KNIME User Training

... and putting all those parts together, you get this final workflow.

Data Reading

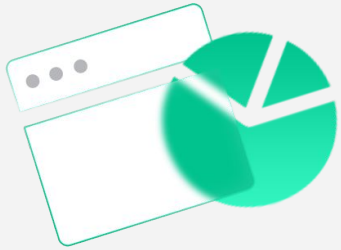


Visualization

Data Manipulation and Aggregation

Training Predictive Models

Data Export and Reporting



Terima Kasih

