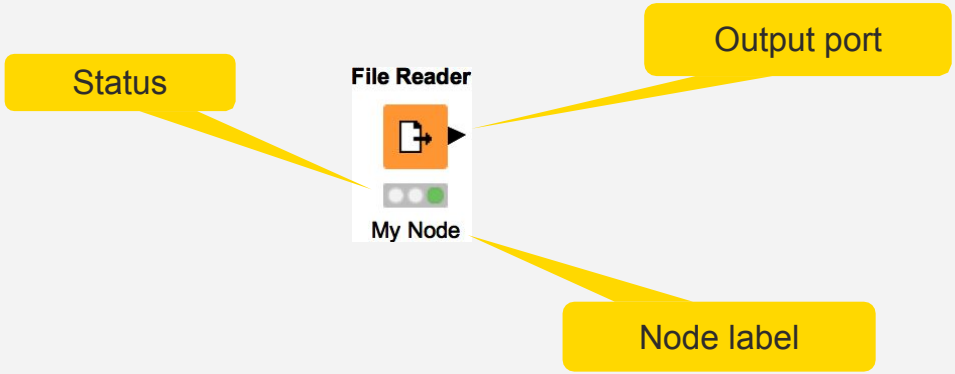
Data Access and Database

Importing Data
Accessing Files

Data Source Nodes

Typically characterized by:

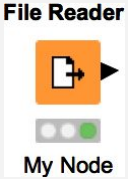
- Orange color
- No input ports, 1-2 output ports



New Node: File Reader

Workhorse of the KNIME Source nodes

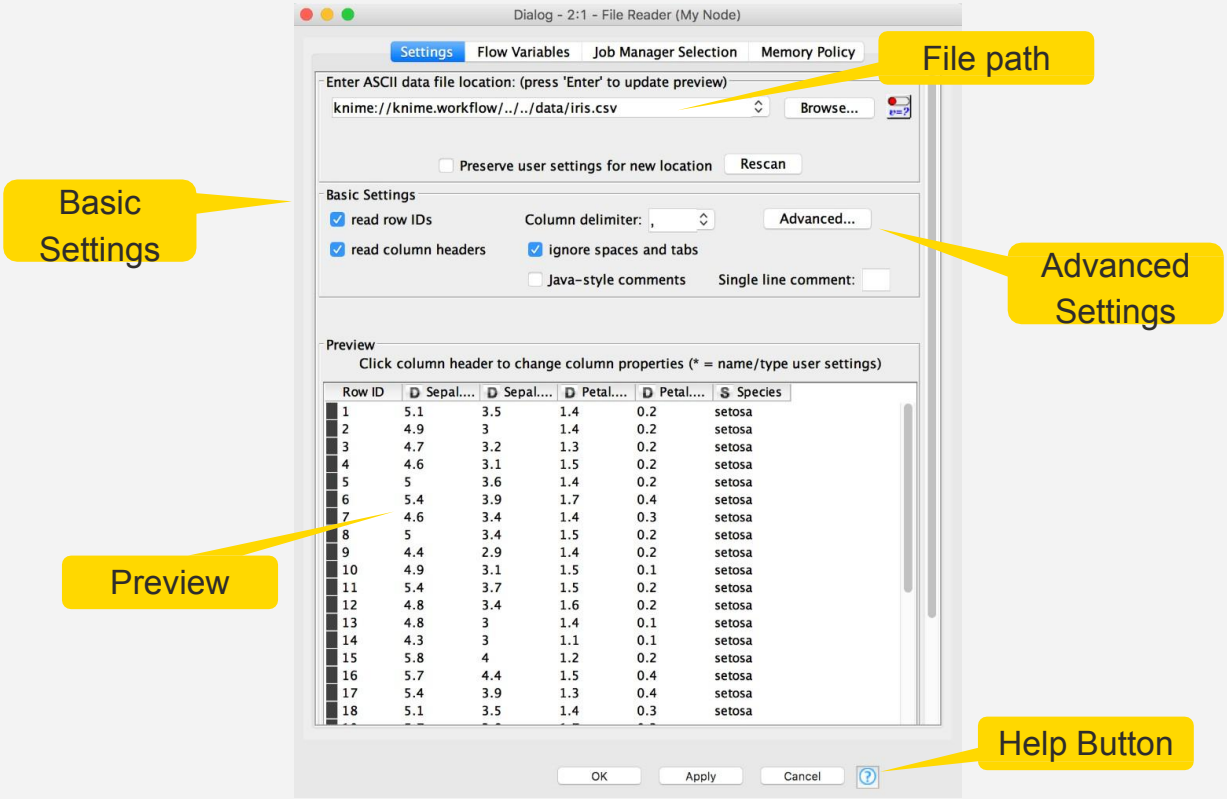
- Reads all text based files (e.g. csv, txt, etc.)
- Many advanced features allow it to read most ‘weird’ files
  - Short lines, inline comments, headers and special encoding



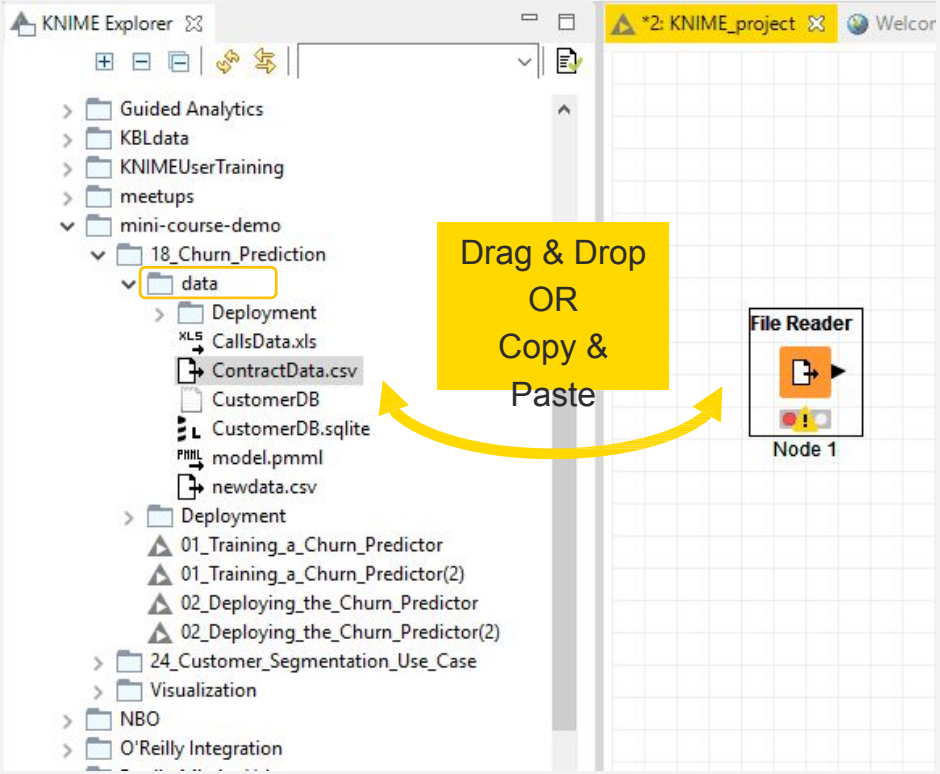
YouTube KNIME TV Channel video:

https://youtu.be/flaHQw-Qhlg

File Reader Configuration

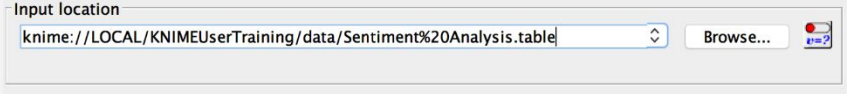


Alternative Faster Way ...

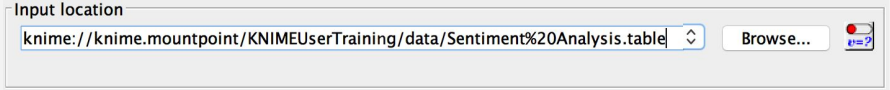


Filenames and the knime:// Protocol

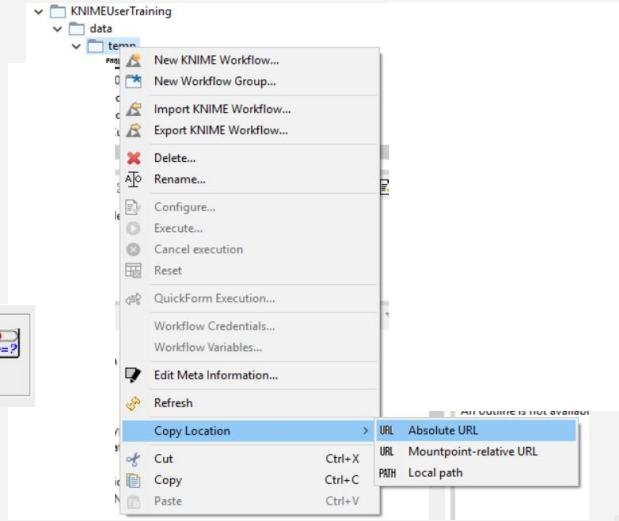
- **Absolute URL**



- **Mountpoint-relative URL**

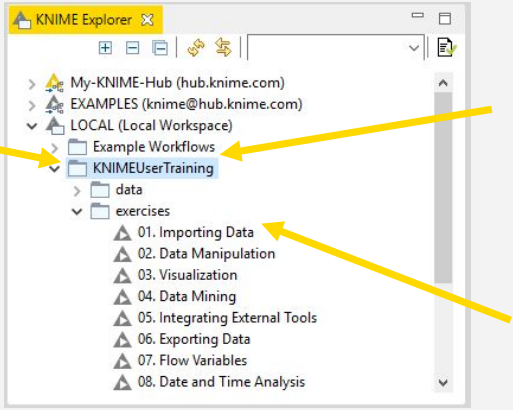


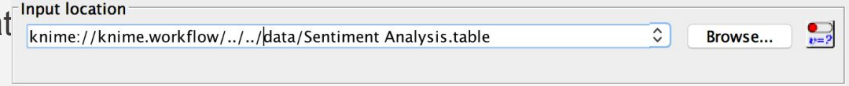
- **Local path**



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

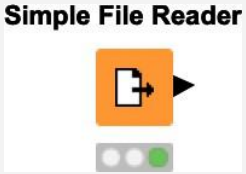


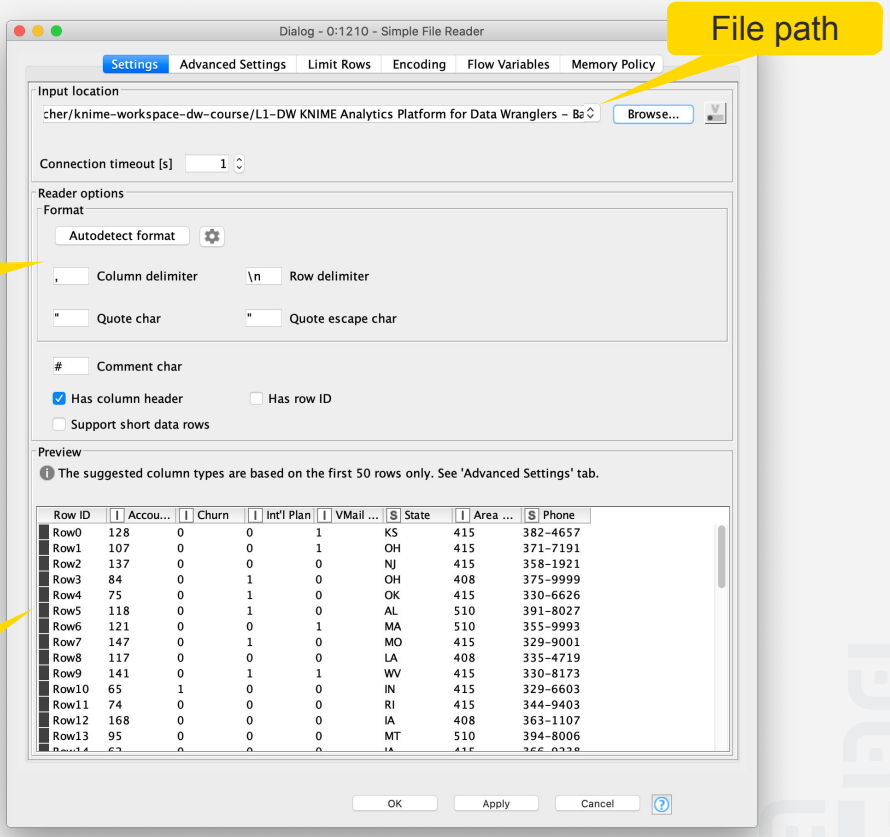


YouTube KNIME TV Channel:
https://youtu.be/U9sP4g4yGwY

New Node: Simple File Reader

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

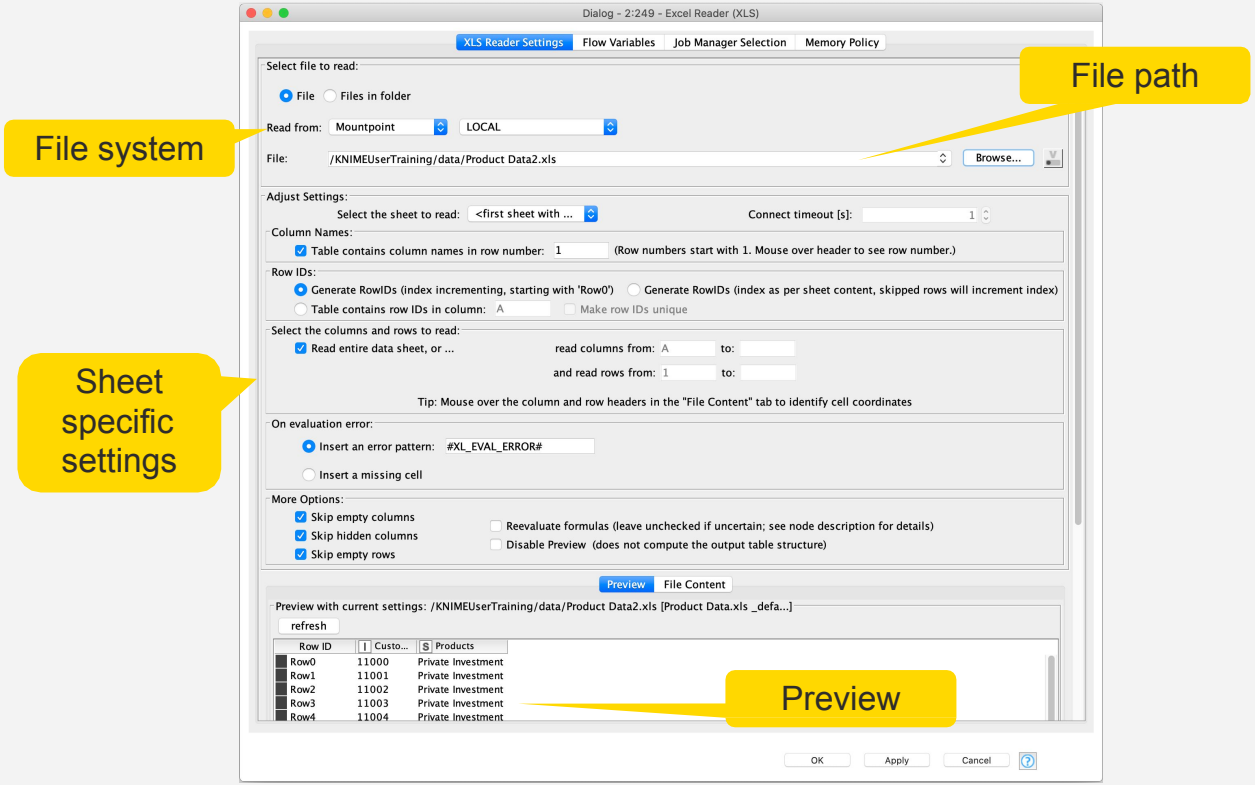




New Node: Excel Reader (XLS)

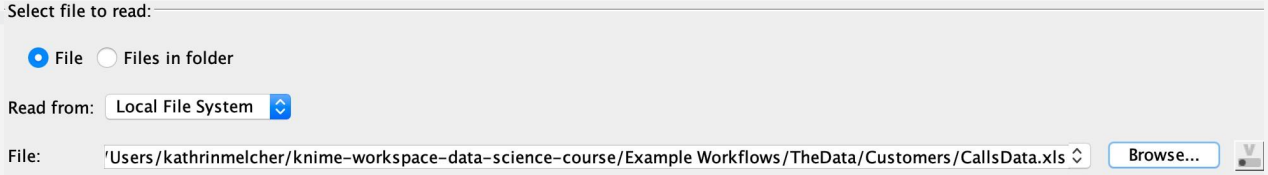
- Reads .xls and .xlsx file from Microsoft Excel
- Supports reading from multiple sheets

Excel Reader Configuration

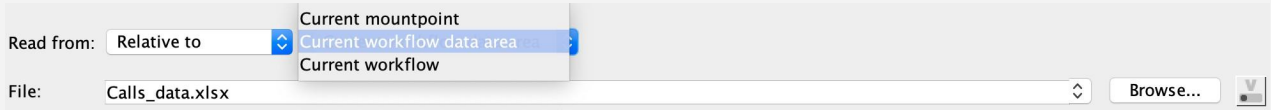


Filenames and the knime:// Protocol

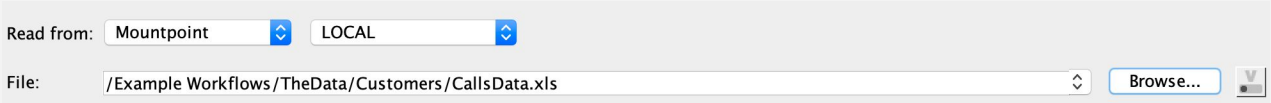
Local File System



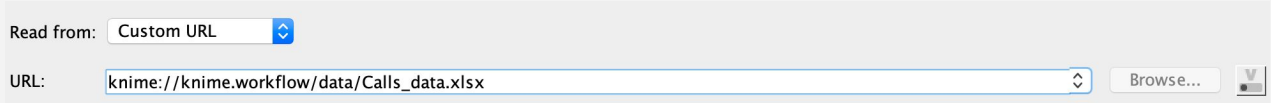
Relative to …



Mountpoint

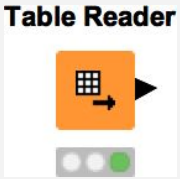


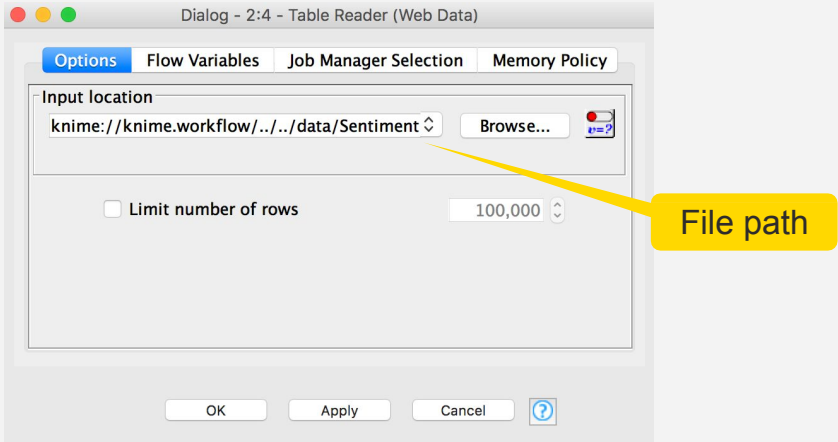
Custom URL



New Node: Table Reader

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

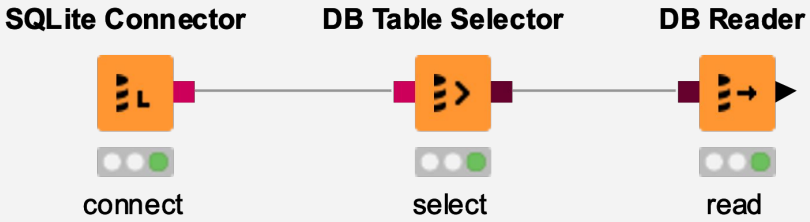




YouTube KNIME TV channel video:
https://youtu.be/tid1qi2HAOo

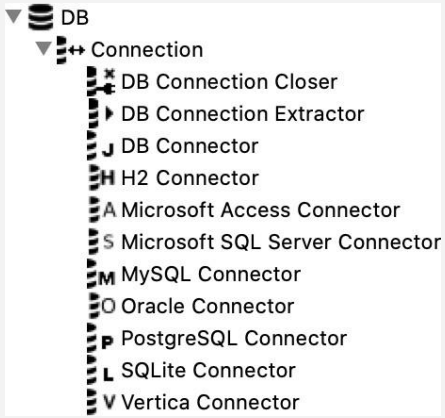
Database Connectivity

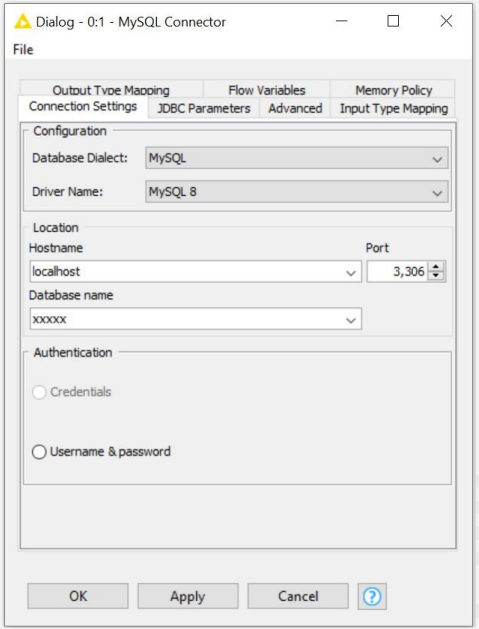
- Read data from any JDBC enabled database
- Write your own SQL or model it using dedicated nodes



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

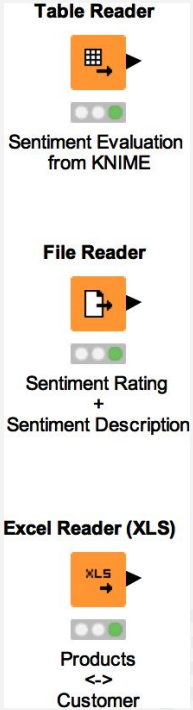
Start with exercise: *Importing Data*

Read the following files

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

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

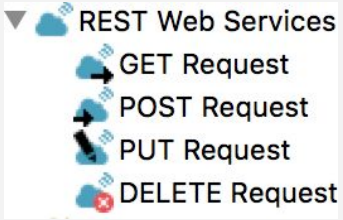
(hint: drag and drop the files from the KNIME Explorer panel to get started)



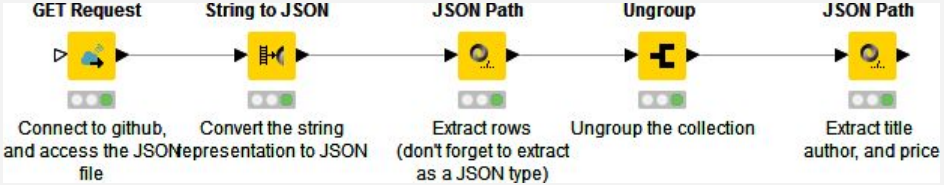
You can download the training workflows from the KNIME Hub:
https://hub.knime.com/knime/spaces/Education/latest/Courses/

RESTful Web Services

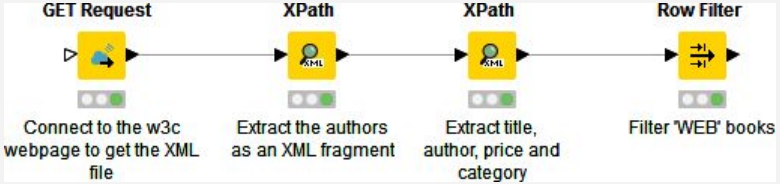
- Use KNIME nodes to interact with RESTful web services
- Send requests using standard HTTP methods



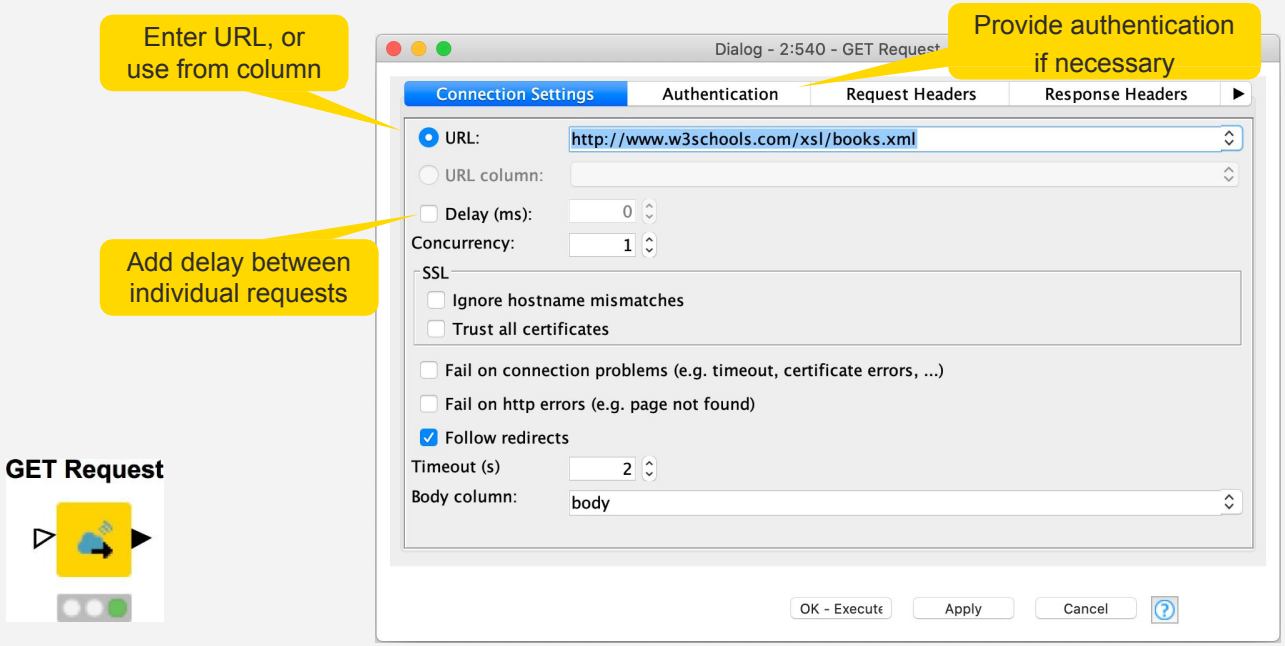
JSON Response:



XML Response:



RESTful Web Services

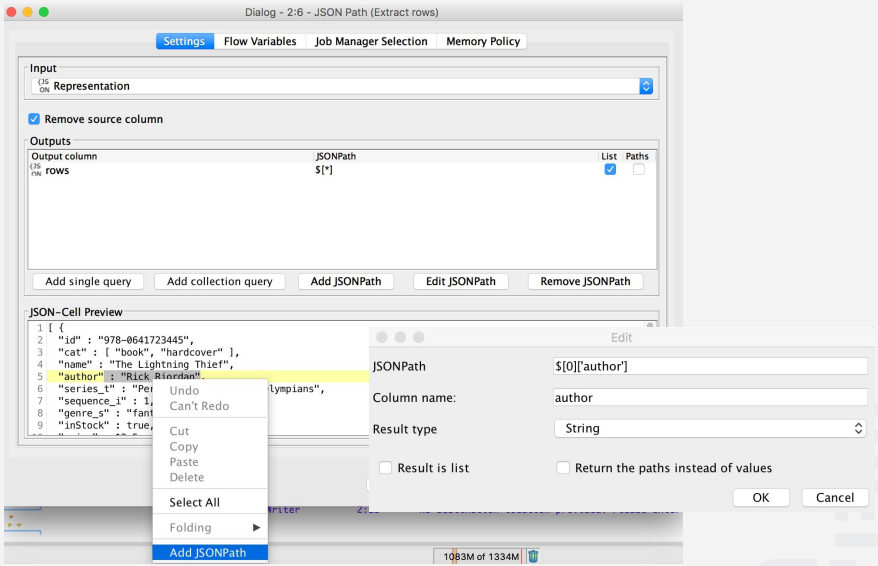


https://www.knime.com/blog/a-restful-way-to-find-and-retrieve-data

https://www.knime.com/blog/OSM-meets-CSV-file-and-Google-API

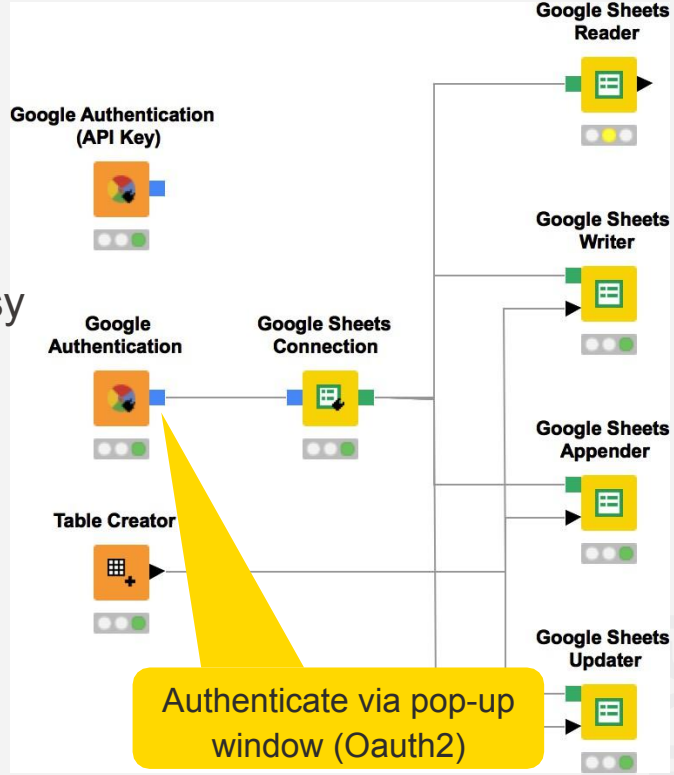
JSON Reader and JSON Path nodes

- 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)



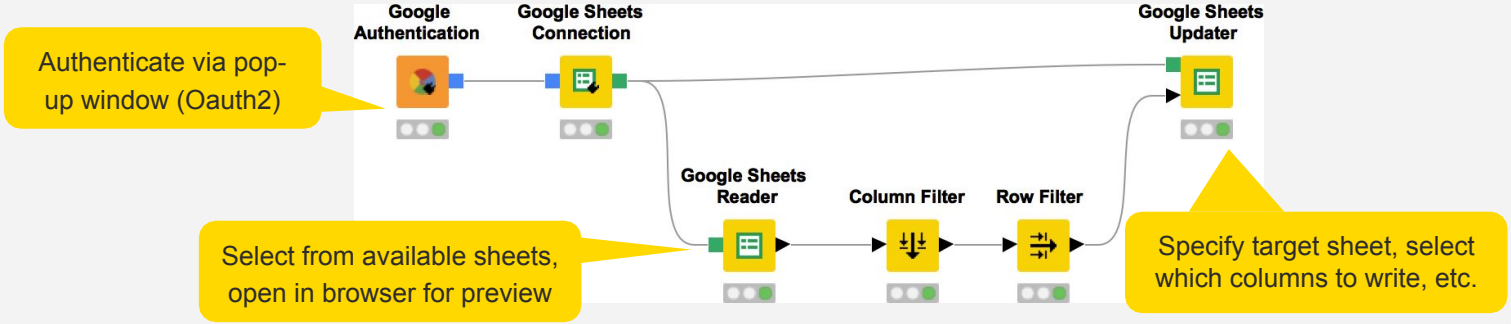
Google Sheets

- 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

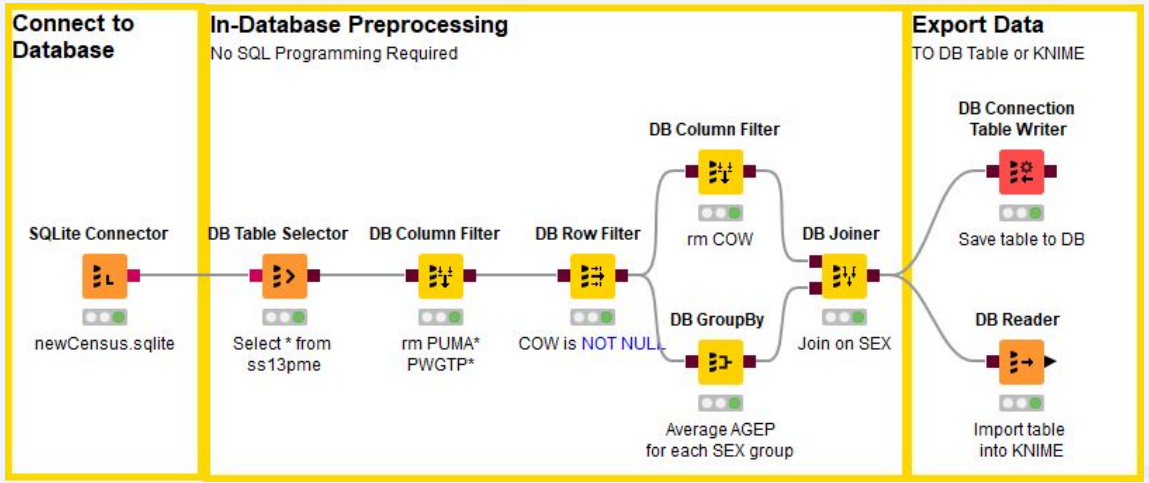
- Select from available sheets on Google Drive
- Transform data in KNIME, or enrich with new data
- Create new sheet or update existing sheets
  - Allows to read from / write to specific range of sheet (e.g. A1:G10)



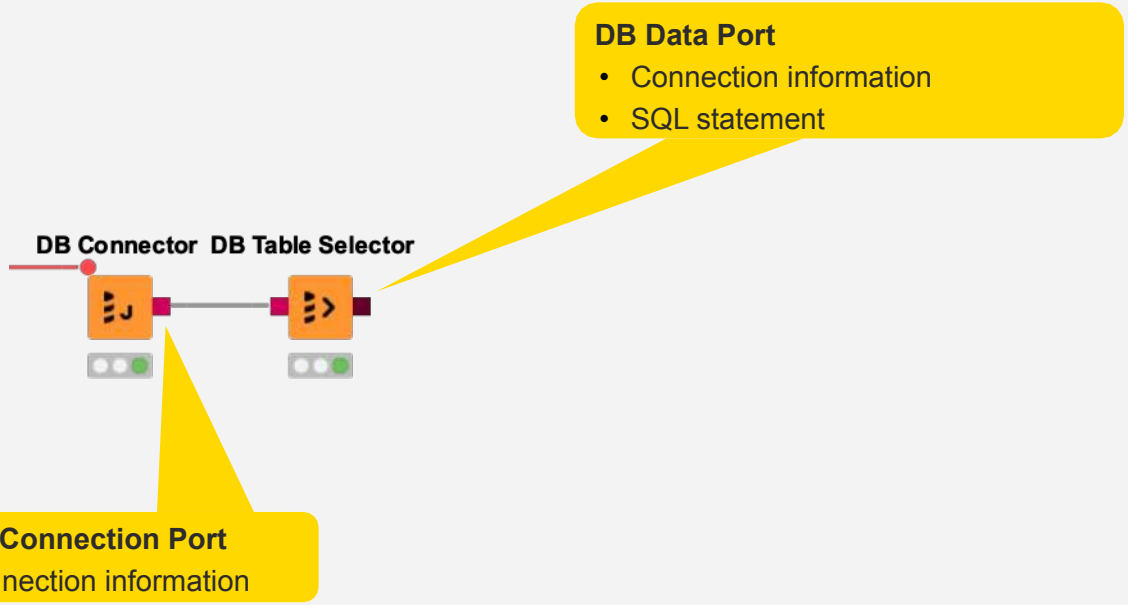
Databases

Database Extension

- Visually 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

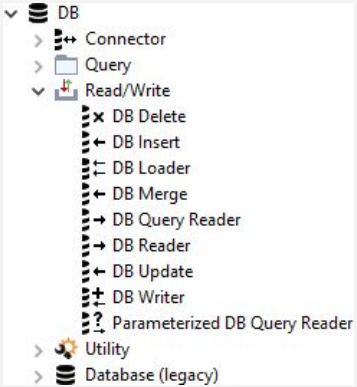


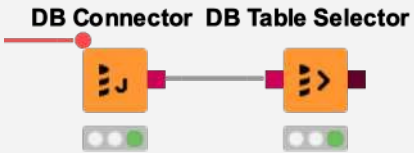
Database Port Types

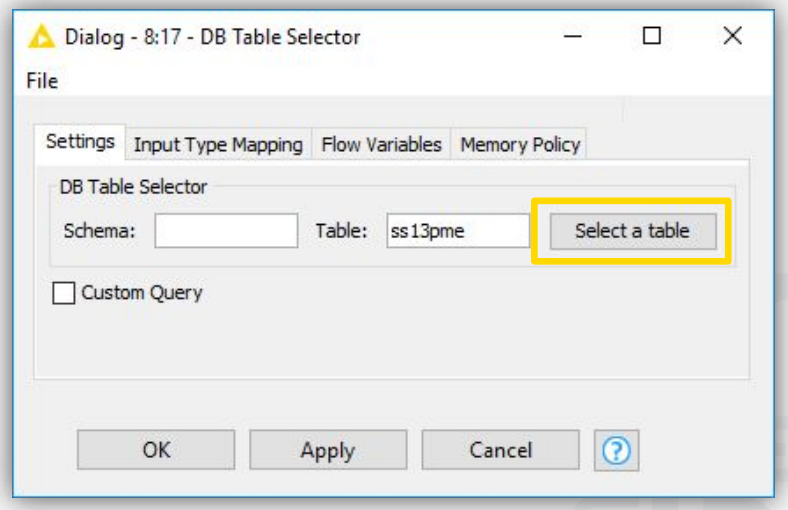


DB Table Selector

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

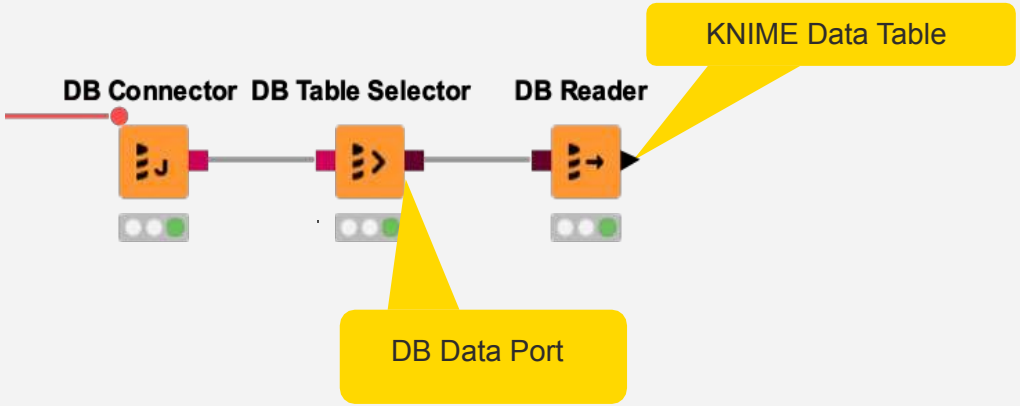






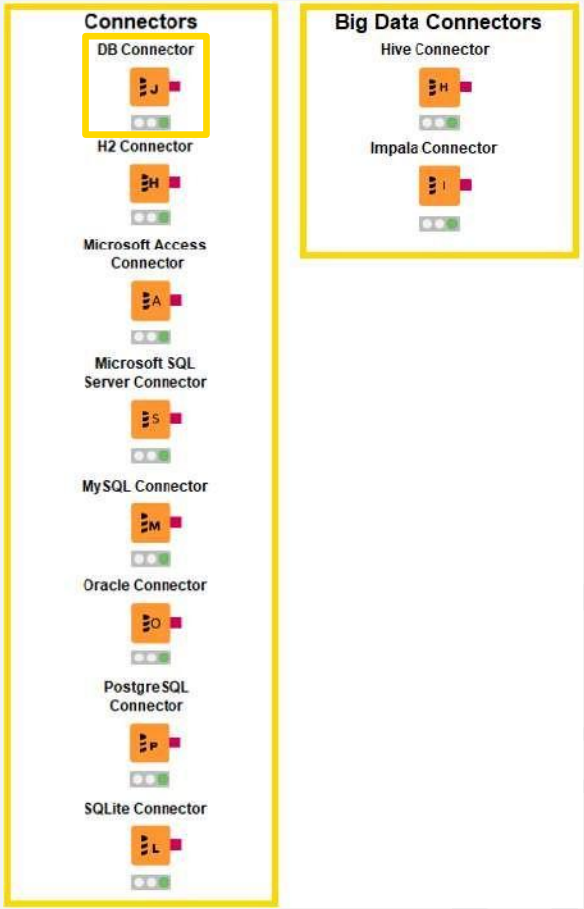
DB Reader

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



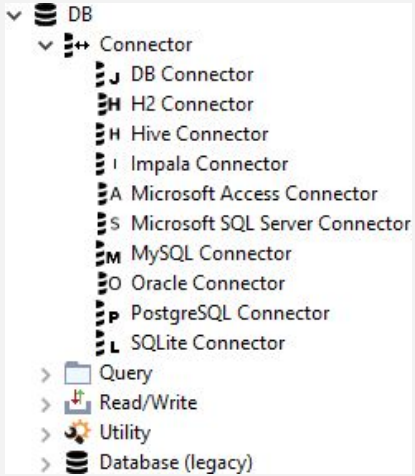
Database Connectors

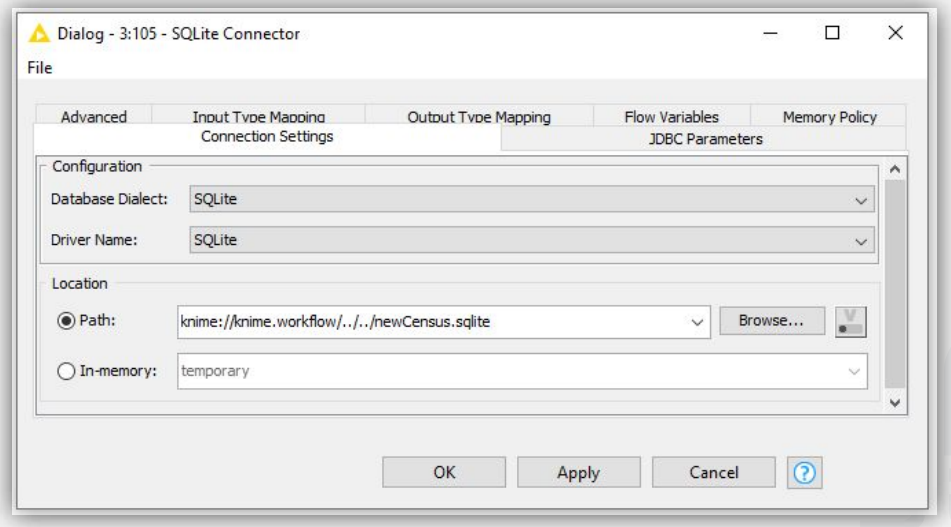
- Dedicated nodes to connect to specific Databases
  - Necessary JDBC driver included
  - Easy to use
  - Import DB specific behavior/capability
- Hive, Impala connectors part of the KNIME Big Data Connectors extension
- General DB Connector
  - Can connect to any JDBC source
  - Register new JDBC driver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



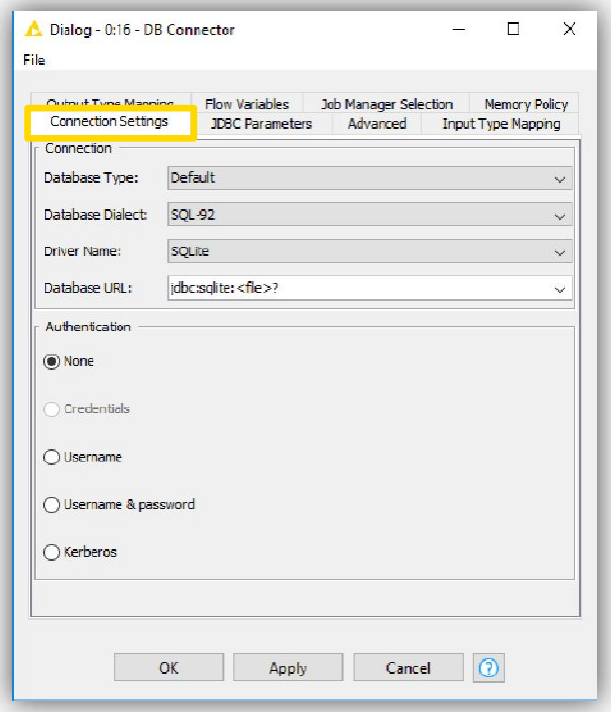
Dedicated Database Connectors

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

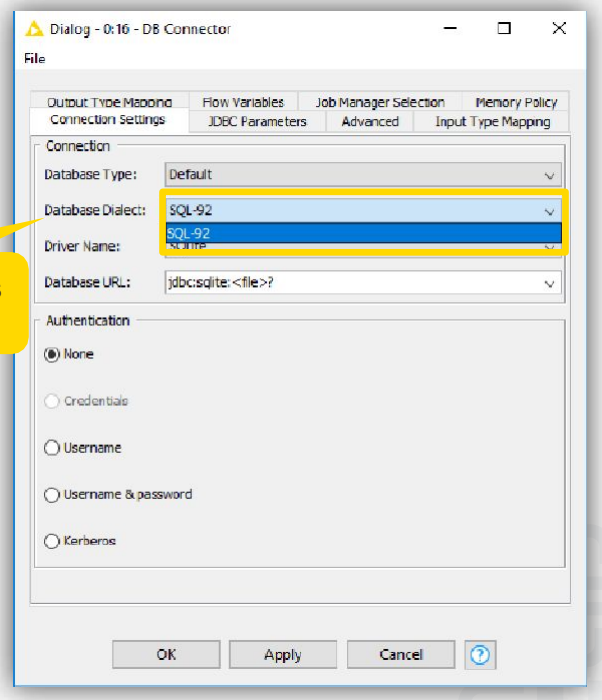




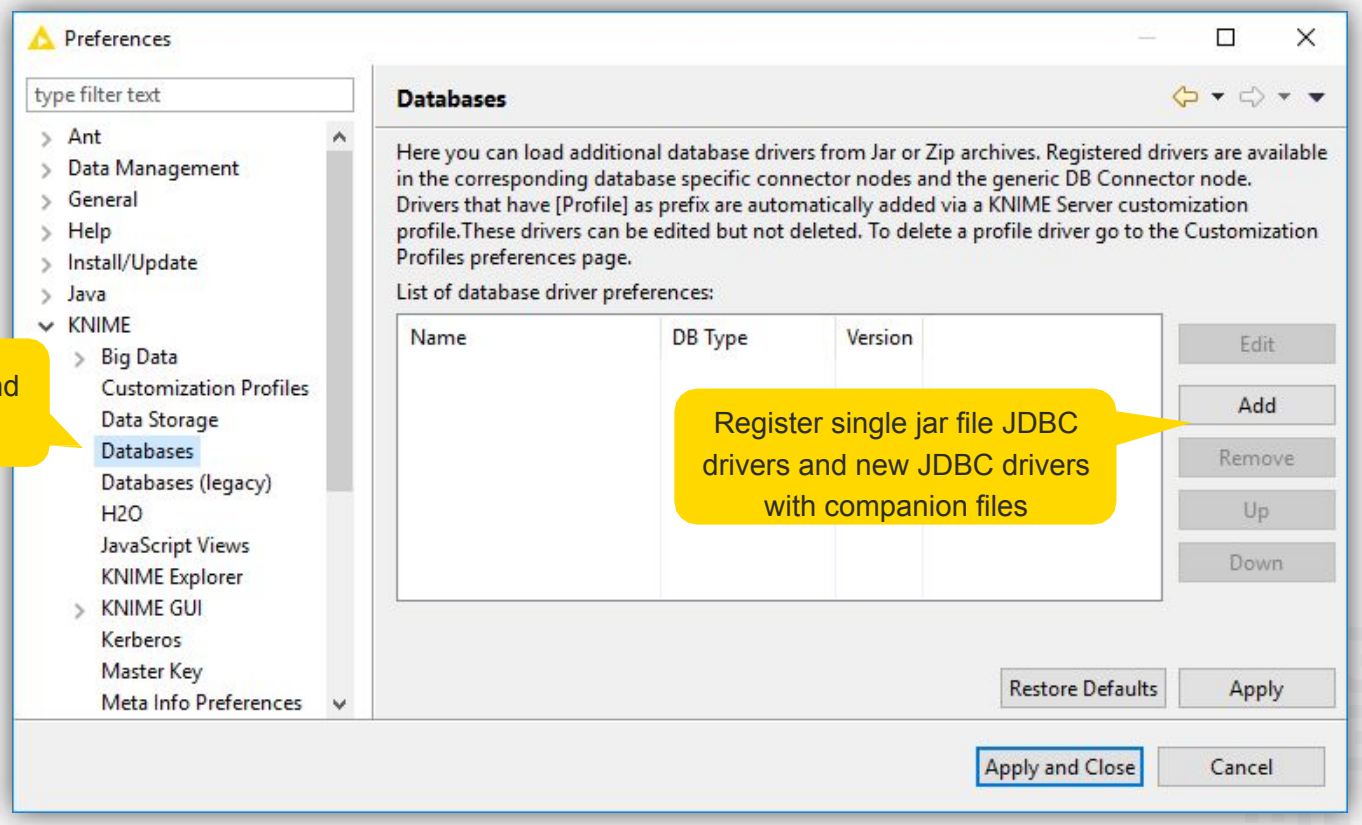
«General» DB Connector Node



Database type defines SQL dialect



Register JDBC Driver



Open KNIME and go to File -> Preferences

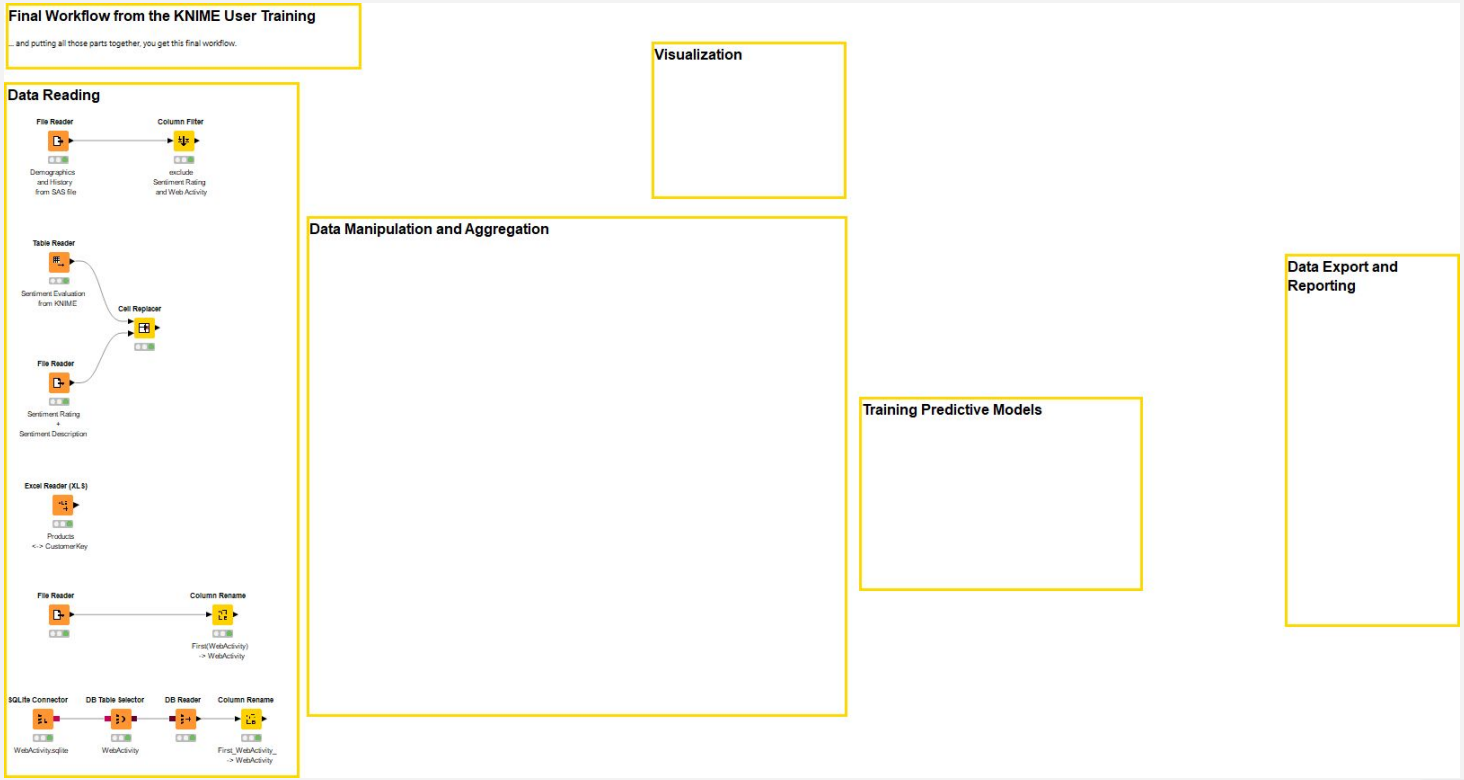
Register single jar file JDBC drivers and new JDBC drivers with companion files

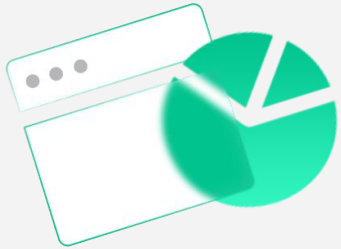
Other Useful Data Sources

- KNIME Analytics Platform provides many more optio 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 Wel

  - Find out more in by downloading the free book “Will they blend” https://www.knime.com/knimepress/download-will-they-blend

Today's Example





Terima Kasih