

Cheat Sheet: Alteryx to KNIME

GENERAL DIFFERENCES

	ALTERYX	KNIME
The basic computation unit	Tool	Node
Execution	Full execution. Only execution of the full workflow is possible. You need tools like Control Parameter to stop/pause execution and inspect results.	Step by step execution. You can execute the workflow fully or node by node. You can inspect the node output after each node's execution.
Parameterization	Control Containers	Flow Variables
Collecting functionalities in one single element	Macros; More specifically the Standard Macros.	Components; Include a workflow segment, can have its own configuration dialog, and its own interactive view.
Repeating a workflow segment	Batch & Iterative Macros	Looping. A number of different loops can be implemented in KNIME using the loop nodes.
Pricing as of April 2024	Designer & Intelligence Suite: For paying KNIME Analytics Platform: Free & open source. KNIME Business Hub: For paying	KNIME Analytics Platform: Free & open source. KNIME Business Hub: For paying
Comments & Descriptions	Comment Tool	Annotation Box
Repository	Tool Palette	Node Repository
Complex interactive visualization framework	Analytic Application	Data App. An interactive visualization framework from the composite view of a component.

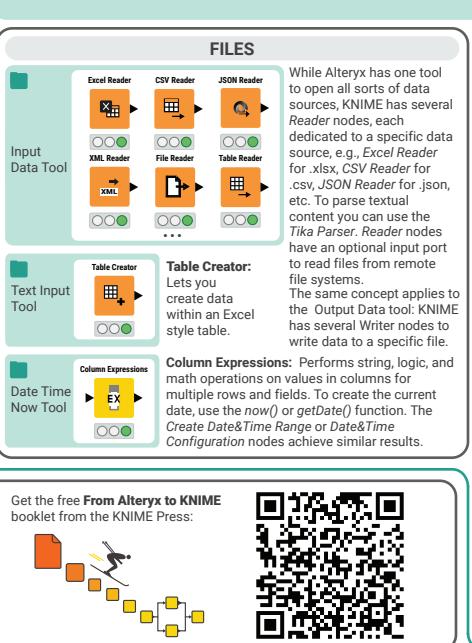
PARSE

Date/Time Tool	   
RegEx Tool	 
Text to Columns Tool	 

REPORTING

Render Tool	 
-------------	---

FILES



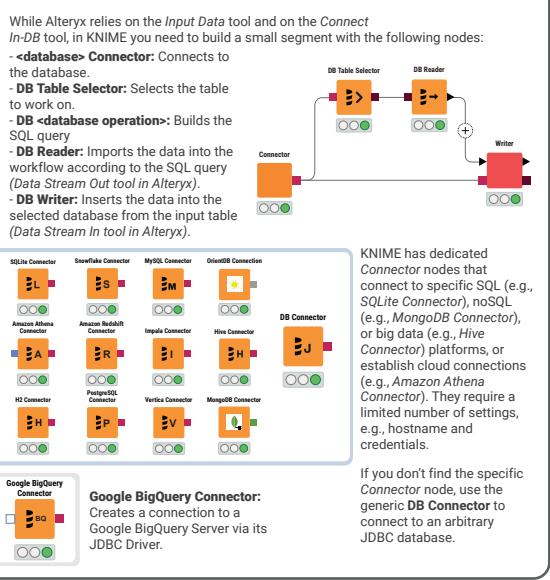
While Alteryx has one tool to open all sorts of data sources, KNIME has several Reader nodes, each dedicated to a specific data source, e.g., Excel Reader for .xlsx, CSV Reader for .csv, JSON Reader for .json, etc. To parse textual content you can use the Tika Parser. Reader nodes have an optional input port to read files from remote file systems. The same concept applies to the Output Data tool: KNIME has several Writer nodes to write data to a specific file.

Column Expressions: Performs string, logic, and math operations on values in columns for multiple rows and fields. To create the current date, use the now() or getDate() function. The Create Date&Time Range or Date&Time Configuration nodes achieve similar results.

Get the free **From Alteryx to KNIME** booklet from the KNIME Press: 

IN/OUT

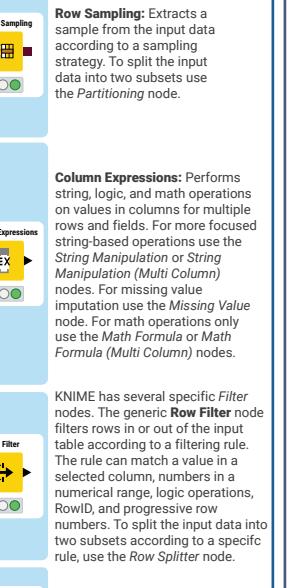
DATABASES



While Alteryx relies on the Input Data tool and on the Connect In-DB tool, in KNIME you need to build a small segment with the following nodes:

- **<database> Connector:** Connects to the database.
- **DB Table Selector:** Selects the table to work on.
- **DB <database operation>:** Builds the SQL query
- **DB Reader:** Imports the data into the workflow according to the SQL query
- **DB Writer:** Inserts the data into the selected database from the input table (Data Stream In tool in Alteryx).

PREPARATION



Row Sampling: Extracts a sample from the input data according to a sampling strategy. To split the input data into two subsets use the Partitioning node.

Column Expressions: Performs string, logic, and math operations on values in columns for multiple rows and fields. For more focused string-based operations use the String Manipulation or String Manipulation (Multi Column) nodes. For missing value imputation use the Missing Value node. For math operations only use the Math Formula or Math Formula (Multi Column) nodes.

KNIME has several specific **Filter** nodes. The generic Row Filter node filters rows in or out of the input table according to a filtering rule. The rule can match a value in a selected column, numbers in a numerical range, logical operations, RowID, and progressive row numbers. To split the input data into two subsets according to a specific rule, use the Row Splitter node.

Missing Value: Define and apply a strategy to replace missing values in the input table - either globally on all columns, or individually for each column separately.

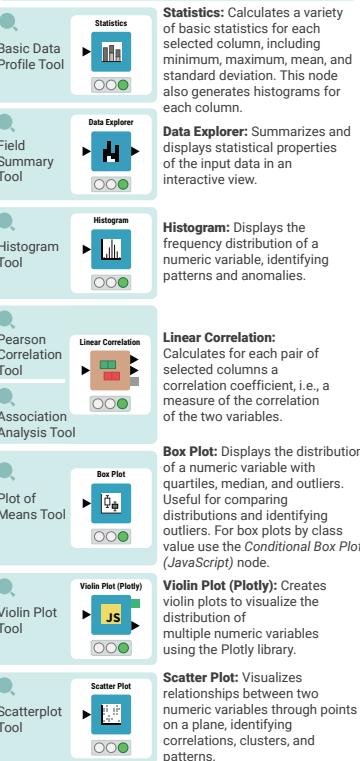
Auto Binner: Allows to group numeric data in intervals (= bins). Use the Numeric Binner node if you want to define custom bins.

Table Manipulator: Performs several column transformations on the input table, e.g., renaming, filtering, re-ordering, and type changing of the input columns. For case-specific operations you can use one of the dedicated nodes like Table Cropper, Column Filter, Column Resorter, Column Renamer, String to Number, etc.

Sorter: Sorts the table in ascending or descending order based on the values of one or more columns. String-compatible columns can be sorted in alphanumeric instead of lexicographic order.

Duplicate Row Filter: Detects duplicate rows and allows removing or keeping them.

DATA INVESTIGATION



Statistics: Calculates a variety of basic statistics for each selected column, including minimum, maximum, mean, and standard deviation. This node also generates histograms for each column.

Data Explorer: Summarizes and displays statistical properties of the input data in an interactive view.

Histogram: Displays the frequency distribution of a numeric variable, identifying patterns and anomalies.

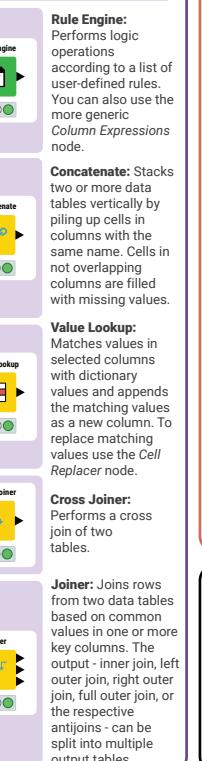
Linear Correlation: Calculates for each pair of selected columns a correlation coefficient, i.e., a measure of the correlation of the two variables.

Box Plot: Displays the distribution of a numeric variable with quartiles, median, and outliers. Useful for comparing distributions and identifying outliers. For box plots by class value use the Conditional Box Plot (JavaScript) node.

Violin Plot (Plotly): Creates violin plots to visualize the distribution of multiple numeric variables using the Plotly library.

Scatterplot: Visualizes relationships between two numeric variables through points on a plane, identifying correlations, clusters, and patterns.

JOIN



Rule Engine: Performs logic operations according to a list of user-defined rules. You can also use the more generic Column Expressions node.

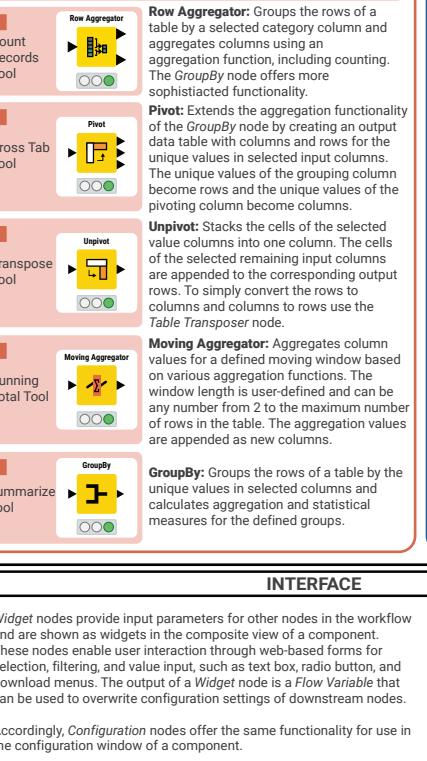
Concatenate: Stacks two or more data tables vertically by piling up cells in columns with the same name. Cells in not overlapping columns are filled with missing values.

Value Lookup: Matches values in selected columns with dictionary values and appends the matching values as a new column. To replace matching values use the Cell Replacer node.

Cross Joiner: Performs a cross join of two tables.

Joiner: Joins rows from two data tables based on common values in one or more key columns. The output - inner join, left outer join, right outer join, full outer join, or the respective antijoins - can be split into multiple output tables.

TRANSFORM



Row Aggregator: Groups the rows of a table by a selected category column and aggregates columns using an aggregation function, including counting. The GroupBy node offers more sophisticated functionality.

Pivot: Extends the aggregation functionality of the GroupBy node by creating an output data table with columns and rows for the unique values in selected input columns. The unique values of the grouping column become rows and the unique values of the pivoting column become columns.

Unpivot: Stacks the cells of the selected value columns into one column. The cells of the selected remaining input columns are appended to the corresponding output rows. To simply convert the rows to columns and columns to rows use the Table Transposer node.

Moving Aggregator: Aggregates column values for a defined moving window based on various aggregation functions. The window length is user-defined and can be any number from 2 to the maximum number of rows in the table. The aggregation values are appended as new columns.

GroupBy: Groups the rows of a table by the unique values in selected columns and calculates aggregation and statistical measures for the defined groups.

INTERFACE

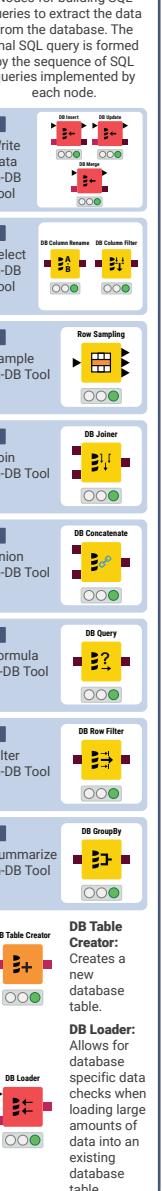
Widget nodes provide input parameters for other nodes in the workflow and are shown as widgets in the composite view of a component. These nodes enable user interaction through web-based forms for selection, filtering, and value input, such as text box, radio button, and download menus. The output of a Widget node is a Flow Variable that can be used to overwrite configuration settings of downstream nodes.

Accordingly, Configuration nodes offer the same functionality for use in the configuration window of a component.

Select:

- Option A
- Option A
- Option B
- Option C

IN-DATABASE



Nodes for building SQL queries to extract the data from the database. The final SQL query is formed by the sequence of SQL queries implemented by each node.

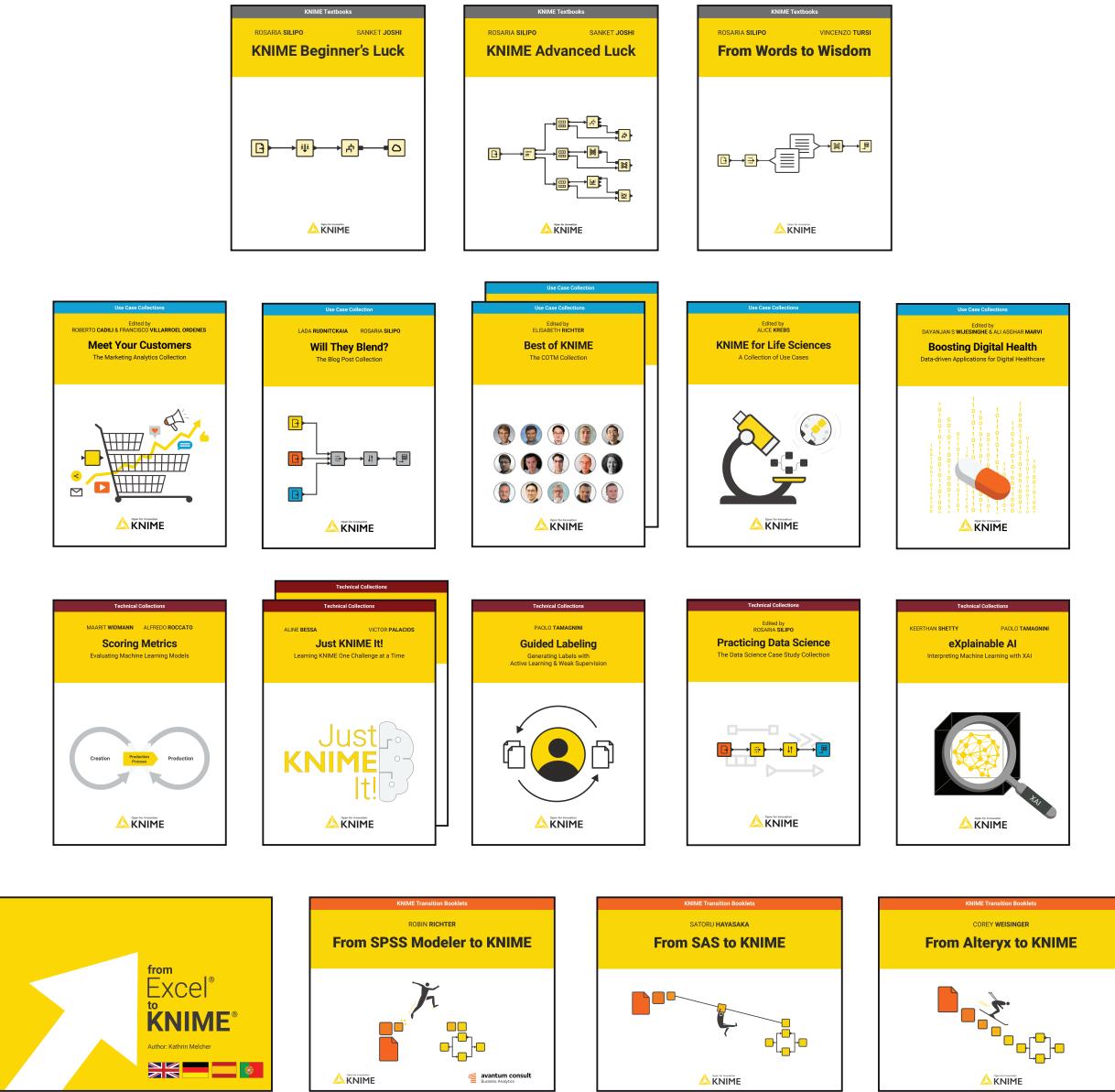
Write Data In-DB Tool
Select In-DB Tool
Row Sampling In-DB Tool
Join In-DB Tool
Union In-DB Tool
DB Query In-DB Tool
DB Row Filter In-DB Tool
DB GroupBy In-DB Tool
DB Table Creator
DB Loader

SPATIAL



KNIME has a full extension of nodes for processing, analyzing, and visualizing Geospatial data, named Geospatial Analytics Extension for KNIME.

Extend your KNIME knowledge with our collection of books from KNIME Press. For beginner and advanced users, through to those interested in specialty topics such as topic detection, data blending, and classic solutions to common use cases using KNIME Analytics Platform - there's something for everyone. Available for download at www.knime.com/knimepress.



Need help?
Contact us!

