TableAggregation Transform

This transform aggregates the source data according a set of key columns. For each value column, an aggregation function has to be specified, such as the sum of a value for all rows with identical key columns.

Aggregation Functions

The following aggregation functions are available for transforms TableAggregation, TableNormalization, and TableDenormalization:

sum	Sum
min	Minimum
max	Maximum
avg	Average
first	The first data entry for one target position
last	The last data entry for one target position
count	Number of data entries written on one target position
count_distinct	Number of distinct data entries written on one target position
group_concat	Concatenates all data entries with separator ","
concat_distinct	Concatenates all distinct data entries with separator "," (i.e. remove duplicate values)
selectivity	Estimates the selectivity (0-100) of a value. The value is defined as (100 * distinctCount / rowCount).
var_samp	The sample variance (square of the sample standard deviation)
stddev_samp	The sample standard deviation
var_pop	The population variance (square of the population standard deviation)
stddev_pop	The population standard deviation
none	No aggregation occurs; the last data set is written (overwriting).

Advanced Settings

Column Include Pattern:	Regular expression for column names to be included as output columns of the transform.
Column Exclude Pattern:	Regular expression for column names to be excluded as output columns of the transform.
Measure Include Pattern:	Regular expression for measure names to be included as output columns of the transform.
Measure Exclude Pattern:	Regular expression for measure names to be excluded as output columns of the transform.
Measure Aggregation:	The aggregation function used for all measures which are added dynamically via the Measure include or exclude pattern.
Keep Source Order:	If set, the order of output rows respects the order in the source with first appearance of a particular key. If not set the order is arbitrary but the performance might be better.
Use caching:	When checked (true), caching will be used; when not checked (false), caching will not be used. See Caching in Extracts and Transforms.

Special Case: Distinct Rows

The distinct rows of a source, i.e. removing of all duplicate rows, can be achieved with this transform by setting "Column include pattern" to "." (i.e. the regular expression for accept all) and no measures. In order to maintain the order of rows the option "keep source order" has to be set.