Grouping

The **Grouping** node performs operations equivalent to a SQL `GROUP BY` clause. Use this component to group records of selected fields and calculate statistical indicators (sum, mean, minimum, and other) for the remaining fields. The component returns one row per group, and the aggregation functions apply calculations to each group individually, not the entire dataset.

▼ Example:

Source table:

Date	Product	Weight, kg
07.12.2015	Apples	20
07.12.2015	Apples	18
07.12.2015	Tomatoes	24
07.13.2015	Tomatoes	22
07.13.2015	Pears	12
07.13.2015	Pears	16

In this example, **Date** and **Product** have been selected as grouping fields, and **Weight, kg** as the parameter field for aggregation. Three aggregation functions have been applied: **Sum**, **Count** (the number of records), and **Mean**.

Result table:

Date	Product	Weight, kg (Sum)	Weight, kg (Count)	Weight, kg (Mean)
07.12.2015	Tomatoes	24	1	24
07.12.2015	Apples	38	2	19
07.13.2015	Pears	28	2	14
07.13.2015	Tomatoes	22	1	22

> As the example shows, the system creates each group from a unique combination of values in the selected grouping fields.

Ports

Input

- **Input data source**: Accepts an input dataset.

Output

- **Output dataset**: Provides the grouped table.

Configuration wizard

The wizard window contains two panes:

- **Available fields**: Lists the fields of the input dataset.
- **Selected fields**: This pane provides two lists: **Group** and **Parameters**.
  - **Group**: Grouping fields.
  - **Parameters**: Fields for which the system will calculate aggregation functions.

▼ Parameter aggregation and supported data types

Aggregation type	ab	12	9.0	0/1	31	∞
Sum		•	•			
Count	•	•	•	•	•	•
Minimum	•	•	•	•	•	•
Maximum	•	•	•	•	•	•
Mean		•	•			
Median		•	•			
Mode	•	•	•	•	•	•
Standard deviation		•	•			
Unique count	•	•	•	•	•	•

Aggregation type	ab	12	9.0	0/1	31	∞
|∅| Null count	•	•	•	•	•	•
1… First	•	•	•	•	•	•
…n Last	•	•	•	•	•	•
1 Only	•	•	•	•	•	•
a,b List	•	•				

Drag fields from the input dataset into the **Group** or **Parameters** lists. Alternatively, use the buttons: **Move to Group** (Alt + G) and **Move to Parameters** (Alt + S). Use the **Filter** field above the **Available fields** list to search for a field by name.

Configure the aggregation method for each parameter in a separate window. Double-click a field in the **Parameters** list or use the context menu to open the selection window. Select the required aggregation methods using the checkboxes; each method writes its result to a separate column.

Below **Aggregation options**, use the **Configure List aggregation option** area to set up the following parameters:

- **Separator**: Set the character that separates values in the list. Select a separator from the list or enter a custom value. Available separators include:
  - Comma (,)
  - Comma with space (,)
  - Semicolon (;)
  - Semicolon with space (;)
  - Space ( )
  - Windows line break (CRLF)
  - Linux line break (LF)
  - Tab ( )
  - No delimiter
- **Text qualifier**: Set this to enclose each list value in specific characters. Select a text qualifier from the list or enter a custom value. Available text qualifiers:
  - Quotes (")
  - Single quotes (')
  - No text qualifier
- **Unique only**: When selected, the result list includes only unique values for the selected field. If a value appears multiple times, the list includes it only once.
- **Sort options**: Select one of the following sort orders for the result list values:
  - No sorting
  - Ascending
  - Descending

The bottom part of the wizard provides two checkboxes:

- **Cache group values**: Select to cache the result data if you want other nodes to reuse it later.
- **Sort result data**: Sort the result table by the grouping fields according to their order in the **Group** list.

Read on: Date and Time Component