

Группировка

Группировка выполняет действия над набором данных, аналогичные действиям SQL-запроса с применением предложения `GROUP BY`. Компонент позволяет объединять записи избранных полей в группы, а для оставшихся полей вычислять статистические показатели (сумму, среднее, минимум и т.д.). Для каждой группы возвращается одна строка. Статистические показатели (или функции агрегации) при этом вычисляются для каждой группы, а не для всего набора в целом.

▼ Пример:

Исходная таблица:

Дата	Товар	Вес, кг
12.07.2015	Яблоки	20
12.07.2015	Яблоки	18
12.07.2015	Помидоры	24
13.07.2015	Помидоры	22
13.07.2015	Груши	12
13.07.2015	Груши	16

В качестве полей-групп выберем поля *Дата* и *Товар*, а поле-параметр (по которому будет проводиться агрегация) — *Вес, кг*. Для примера применим три функции агрегации: сумма, количество записей и среднее.

Результирующая таблица:

Дата	Товар	Вес, кг (Сумма)	Вес, кг (Количество)	Вес, кг (Среднее)
12.07.2015	Помидоры	24	1	24
12.07.2015	Яблоки	38	2	19
13.07.2015	Груши	28	2	14
13.07.2015	Помидоры	22	1	22

Как видно из примера, группа образуется уникальным сочетанием значений полей, выбранных в качестве группировочных.

Вход

-  **Входной источник данных** — порт для подключения входного набора данных.

Выход

-  **Выходной набор данных** — порт отдающий сгруппированную таблицу.

Мастер настройки

Окно мастера поделено на две области.

- Доступные поля** — содержит список полей входного набора данных.
- Выбранные поля** — делится на списки *Группа* и *Показатели*.
 -  **Группа** — поля группировки.
 -  **Показатели** — поля, по которым рассчитываются функции агрегации.

▼ Агрегация показателей и возможные типы данных

Вид агрегации	ab	12	9.0	0/1	31	∞
 Сумма		•	•			
 Количество	•	•	•	•	•	•
 Минимум	•	•	•	•	•	•
 Максимум	•	•	•	•	•	•
 Среднее		•	•			
 Медиана		•	•			
 Мода	•	•	•	•	•	•
 Стандартное отклонение		•	•			
 Количество уникальных	•	•	•	•	•	•
 Количество пропусков	•	•	•	•	•	•
 Первый	•	•	•	•	•	•
 Последний	•	•	•	•	•	•
 Единственный	•	•	•	•	•	•

Вид агрегации	ab	12	9.0	0/1	$\frac{31}{31}$	∞
a,b Список	.	.				

Для настройки требуется переместить поля входного набора в списки *Группа* или *Показатели*, перетаскивая их мышью. Так же это можно сделать с помощью кнопок :  **Переместить в Группу** (комбинация горячих клавиш **Alt+G**) и Σ **Переместить в Показатели** (комбинация горячих клавиш **Alt+S**). Над списком доступных полей расположено поле  *Фильтрация*, оно позволяет найти поле по имени или его части.

Настройка метода агрегации для каждого показателя производится в отдельном окне. Чтобы его открыть, нужно дважды кликнуть по полю в списке *Показатели* или вызвать это окно из контекстного меню. Далее отметить галочками нужные методы агрегации. Результат для каждого метода будет записан в отдельный столбец.

Под *Вариантами агрегации* находится область *Настройки варианта агрегации "Список"*, включающая следующие настройки:

- Разделитель — задаёт символ, которым будут разделены значения в списке. Разделитель можно выбрать из предопределенного списка или ввести свое значение в поле. Список значений разделителя:
 - Запятая (,)
 - Запятая с пробелом (,)
 - Точка с запятой (;)
 - Точка с запятой с пробелом (;)
 - Пробел ()
 - Перенос строки Windows (CRLF)
 - Перенос строки Linux (LF)
 - Табуляция (␣)
 - Без разделителя
- Ограничитель строк — параметр позволяет заключить каждое значение списка в заданные символы. Символ-ограничитель можно выбрать из предлагаемого списка или ввести свое значение. Список значений ограничителя:
 - Кавычки (")
 - Одинарные кавычки (')
 - Без ограничителя строк
- Только уникальные — при выборе данного параметра в результирующем списке отобразятся только уникальные значения выбранного поля (т.е., если в поле какое-то значение встречается несколько раз, то в список будет включено только одно такое значение).
- Параметры сортировки — для значений результирующего списка можно задать один из следующих видов сортировки:
 - Без сортировки
 - По возрастанию
 - По убыванию

В нижней части мастера расположены два параметра:

- **Кэшировать значения групп** — результирующие данные будут закэшированы для использования последующими узлами.
- **Сортировать результирующие данные** — данные в результирующей таблице будут отсортированы по полям группировки в зависимости от их последовательности расположения в списке *Группы*.