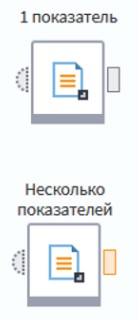
**Рабочая тетрадь №1. Часть 3**

**Настройка портов**

Перенесём в поле построения сценария 2 компонента **Текстовый файл** и импортируем текстовые файлы **1 показатель.txt** и **Несколько показателей.txt**.

# Типы полей:

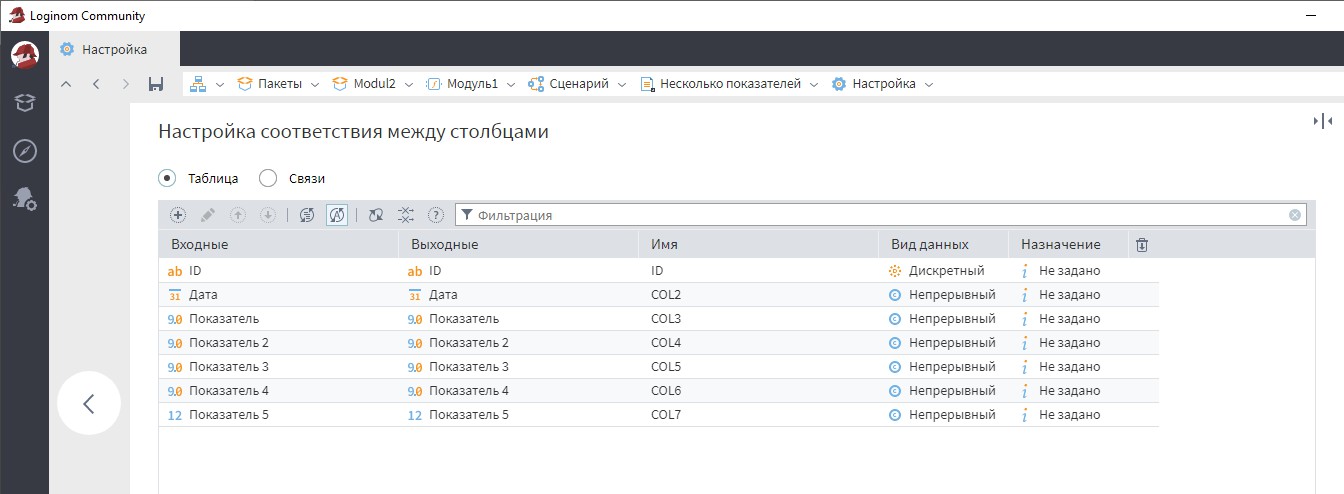
* стоковый (символы);
* логический (истина/ложь/пусто);
* дата/время;
* целый (целые числа);
* вещественный (числа с плавающей точкой);
* переменный (все вышеперечисленные типы данных одновременно).

У поля **Показатель 5** поменяем тип данных на **целый**. **Вид данных поля**:

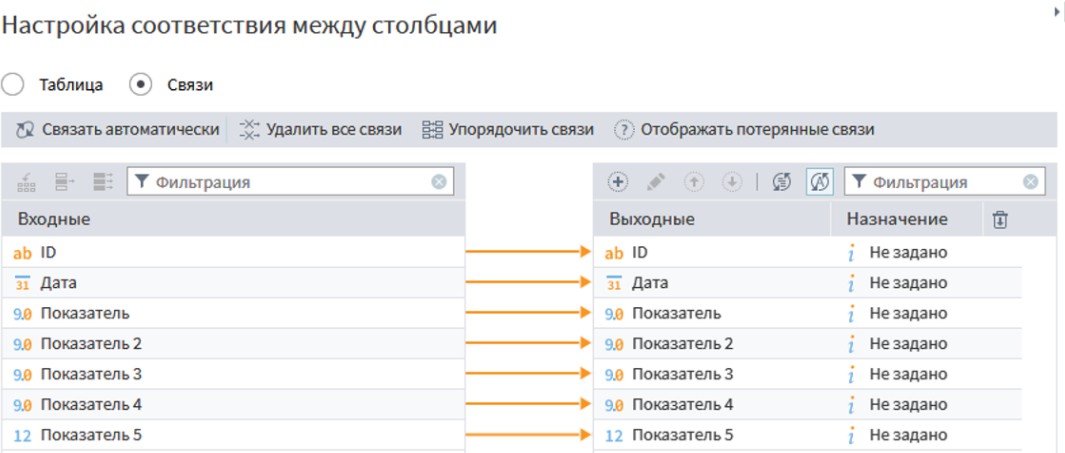
* **непрерывный** (значения в столбце принимают любое значение в рамках своего типа);
* **дискретный** (значения в столбце принимают ограниченное число значений).

**Дискретные** данные – как правило строковые, реже целого и вещественного типов. От вида данных зависит работа фильтра строк, расчет статистики по столбцу и некоторые алгоритмы машинного обучения.

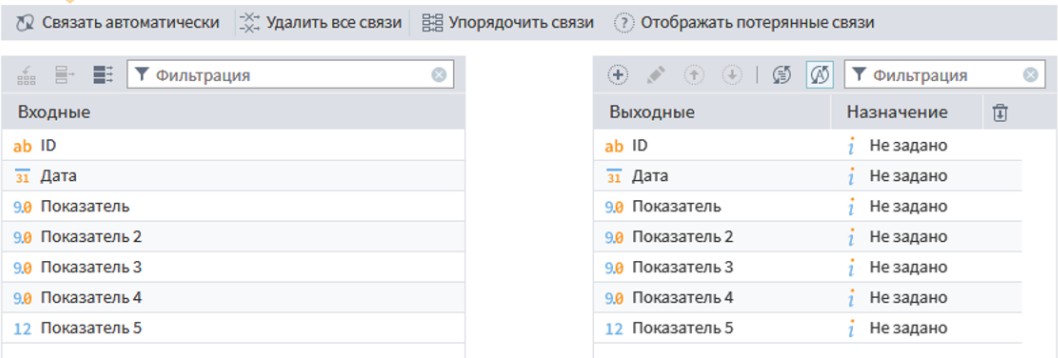
На странице **Настройка соответствия между столбцами** (мэппинг полей) определяется соответствие между наборами входных и выходных полей.



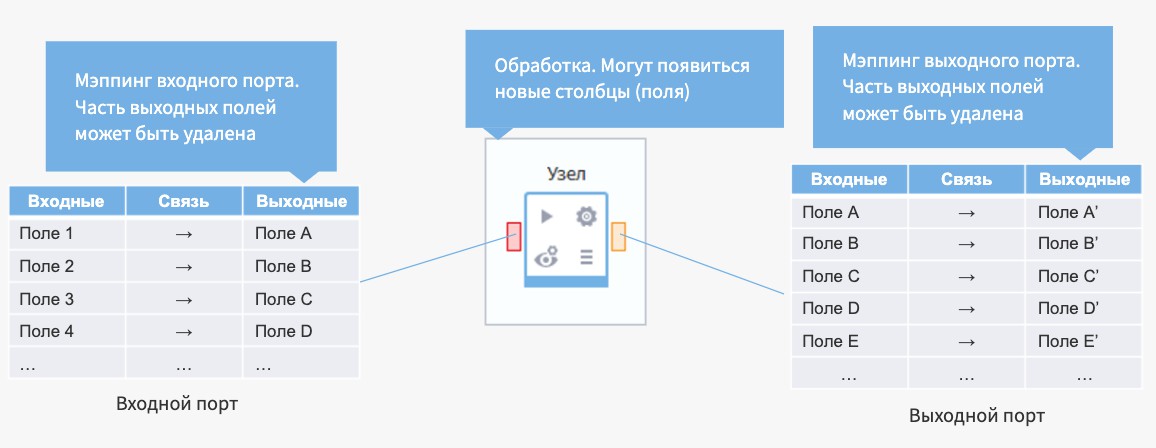
Режимы окна: **Таблица** и **Связи**.



В режиме представления со связями можно удалить все связи входных и выходных полей операции мэппинга с помощью команды: **Удалить все связи** в панели инструментов.



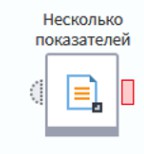
Обратная операция: **Связать автоматически**. Алгоритм связывания пытается связать входные поля, у которых нет связи и выходные поля. Сначала используется пара признаков **Тип данных** – **Имя поля**, далее: **Тип данных** – **Метка**, учитывая совместимость типов данных.

Рассмотрим **мэппинг** подробнее. Допустим, есть узел с одним входным и одним выходным набором данных. К входному порту данного узла идёт связь из другого узла. Набор данных пройдёт два этапа мэппинга: на входе, на выходе.

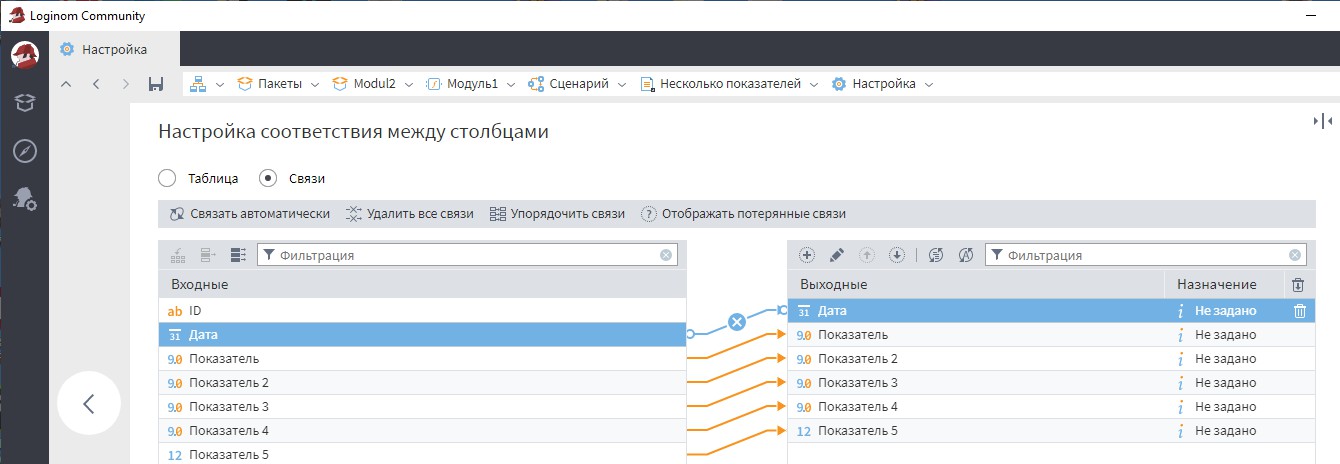
На каждом этапе мэппинга число операций сопоставления равно количеству портов.

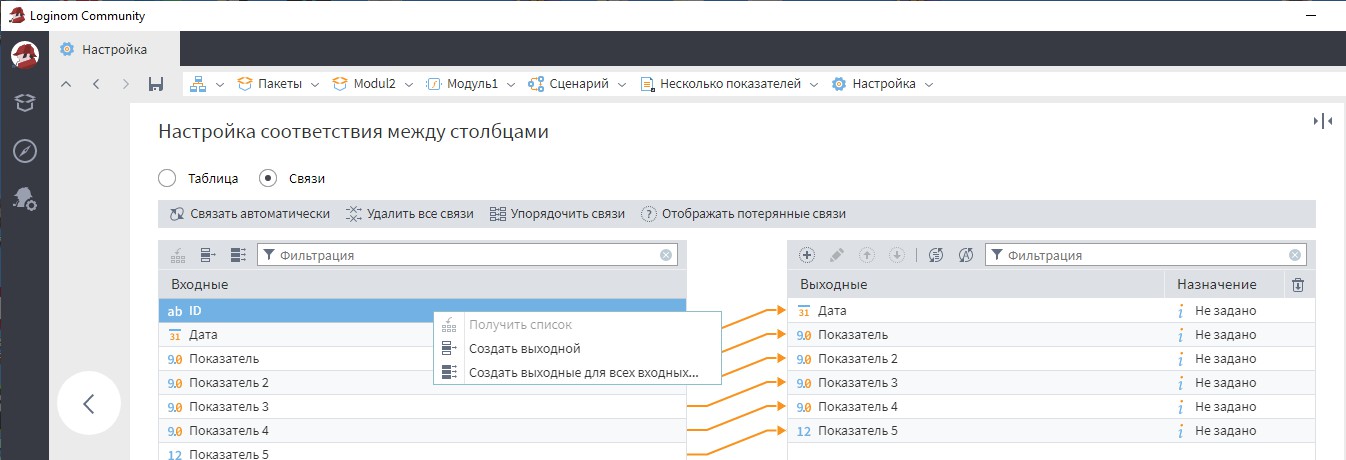
Чтобы удалить одну связь по полю, необходимо нажать на стрелочку связи левой кнопкой мыши, в центре стрелки появится знак удаления, нажав на который связь удалится.



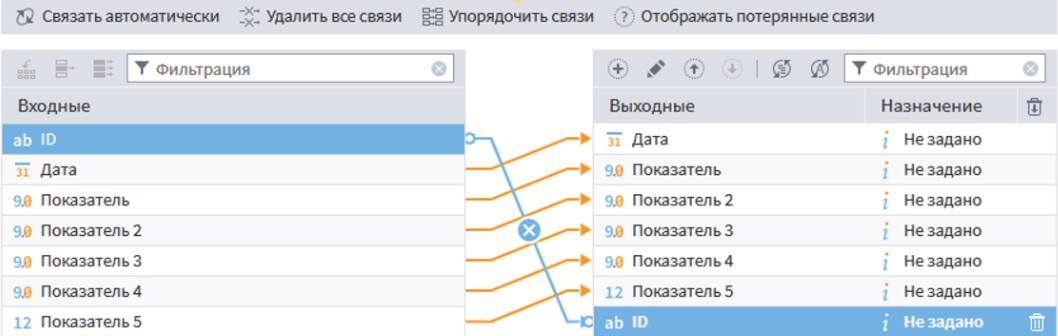
Мэппинг успешный, если у каждого входного поля есть связь с выходным полем. Иначе при попытке перейти на следующий шаг появится предупреждение об ошибке. При этом мастер настройки можно завершить, сохранив настройки. В таком случае выходной порт будет окрашен в красный цвет, что говорит о том, что мэппинг полей прошёл не успешно, выполнить узел не удастся, так как выходной порт не сконфигурирован.

Если удалить не связь, а выходное поле, оно не сможет участвовать в дальнейшей обработке. Для этого нажмем на знак корзины справа от названия поля.



Чтобы восстановить удалённое поле, можно использовать команду **Создать выходной** контекстного меню, вызываемого при нажатии правой кнопки мыши на нужном входном поле.

Появится новое выходное поле, а стрелочка связи будет пересекать все остальные стрелки.



Чтобы не было пересечений связей, можно воспользоваться командой

**Упорядочить связи** на панели инструментов.

Также, восстановить удалённое выходное поле можно с помощью команды

**Добавить столбец (**кнопка  на панели инструментов**)**. Этот способ требует ручного введения имени, метки, типа данных.

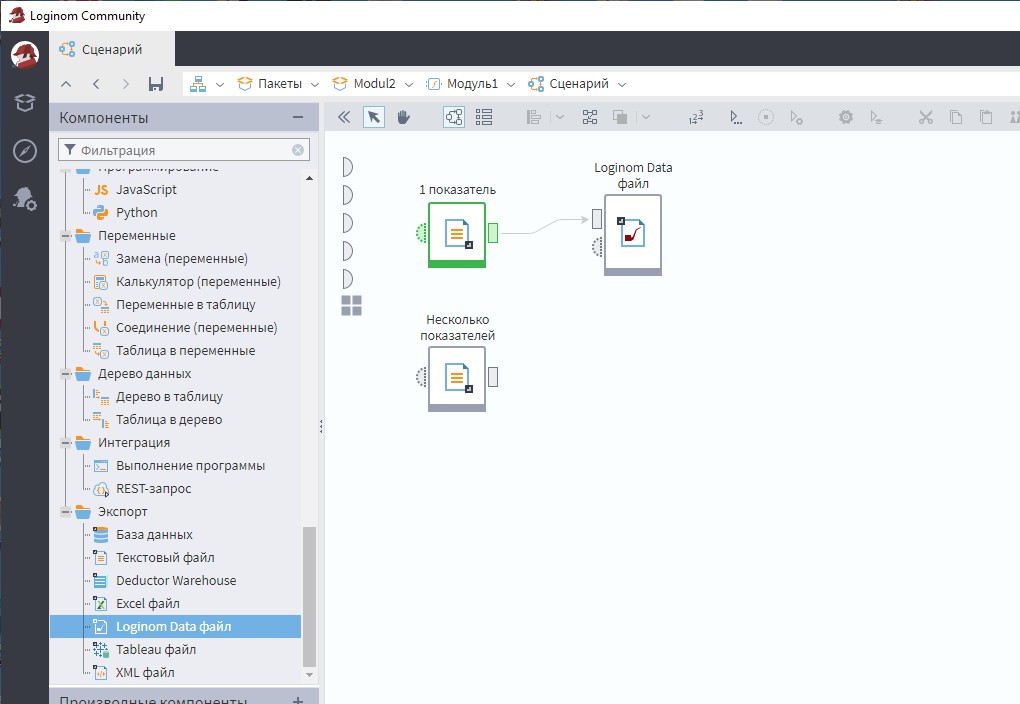
После ручных правок в настройке соответствия, отключается кнопка автосинхронизации  в панели инструментов. Необходимо её включить.

# Практическое задание №9

Создать сценарий импорта файла формата **txt** и экспорта его в формат **lgd**.

А. Добавим в поле сценария компонент экспорта **Loginom Data файл**. Экспортируем с помощью него набор данных **1 показатель** в файл формата **lgd**. Для этого:

* Создадим связь между выходным портом узла **1 показатель** и входным портом **Loginom Data файл**.
* Активируем узел **1 показатель**.

Если входной порт **Loginom Data файл** серого цвета (не красного), значит мэппинг полей прошёл успешно, порт сконфигурирован, произошла синхронизация полей.

**Синхронизация полей** – однократный запуск последовательности действий с целью мэппинга полей без участия пользователя.

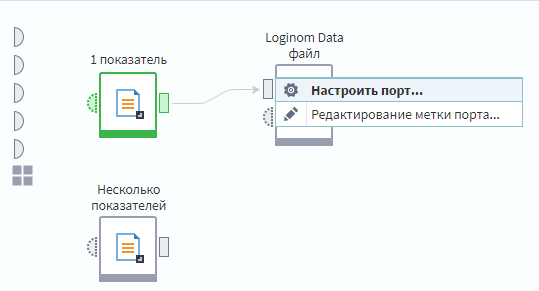
# Алгоритм синхронизации полей:

1. Попытка автоматически связать входные и выходные поля.
2. Добавление полей в список выходных, основываясь на списке входных полей, автоматическое связывание входных и выходных полей.
3. Обновление списка выходных полей.

Когда список полей входного порта пуст (во время первого мэппинга), выполняется только п.2.

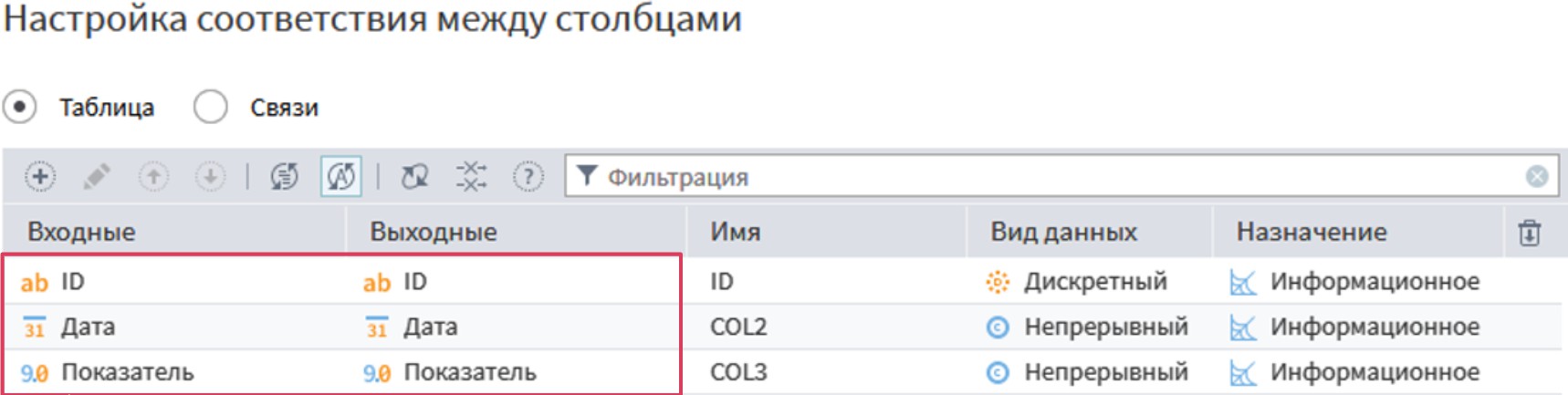
Синхронизация проходит в следующих случаях:

* + если вызвать синхронизацию командой **Синхронизация** в мастере настройки порта;
  + если провести связь портов двух узлов, когда пуст список полей входного порта;
  + в порте включена автосинхронизация, узел активируется;
  + открывается страница мастера настройки соответствия столбцов при включенной автосинхронизации порта.

Войдём в настройки входного порта узла **Loginom Data файл** нажатием по порту правой кнопкой мыши и выбрав команду **Настроить порт**.

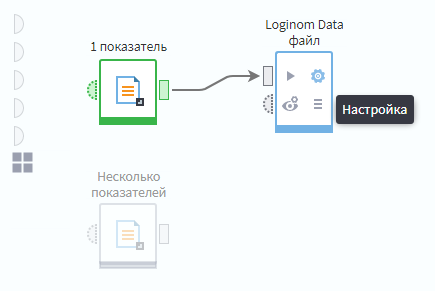
Можно убедиться в том, что синхронизация прошла успешно: появилось 3

выходных поля, идентичные входным, между ними есть связь.

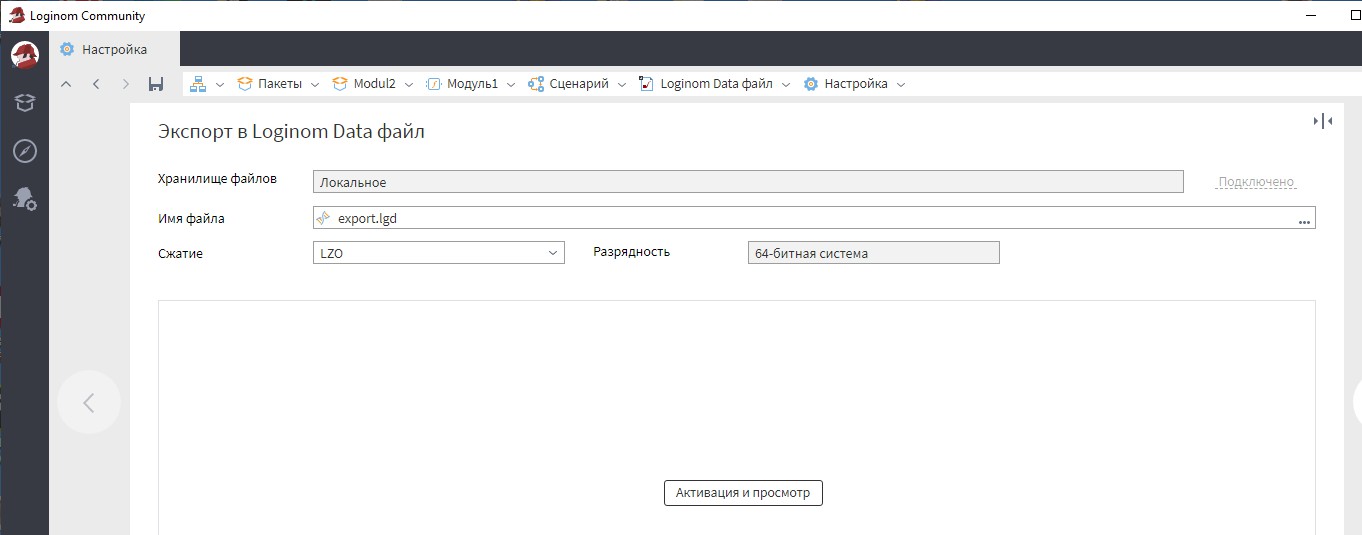


Включена кнопка автосинхронизации, значит, если изменится набор полей, поступающих на входной порт, список выходных полей будет автоматически синхронизирован.

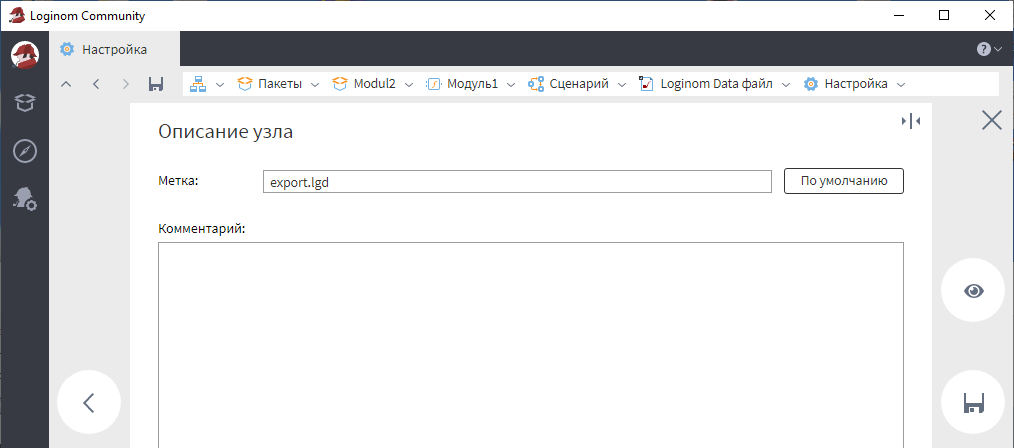
Запустим мастер настройки узла.



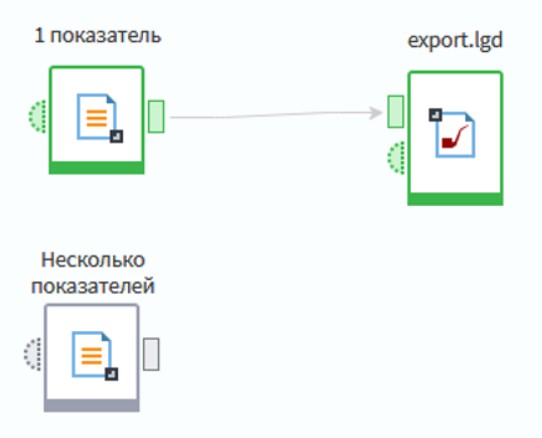
Появится окно **Экспорт в Loginom Data файл**. В поле Имя файла необходимо ввести название файла и расширение **lgd.**



В следующем окне (**Описание узла**) в поле **Метка** также напишем название нашего файла.



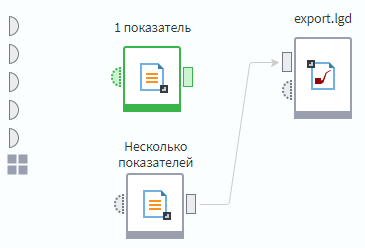
Сохраним изменения и выполним узел.



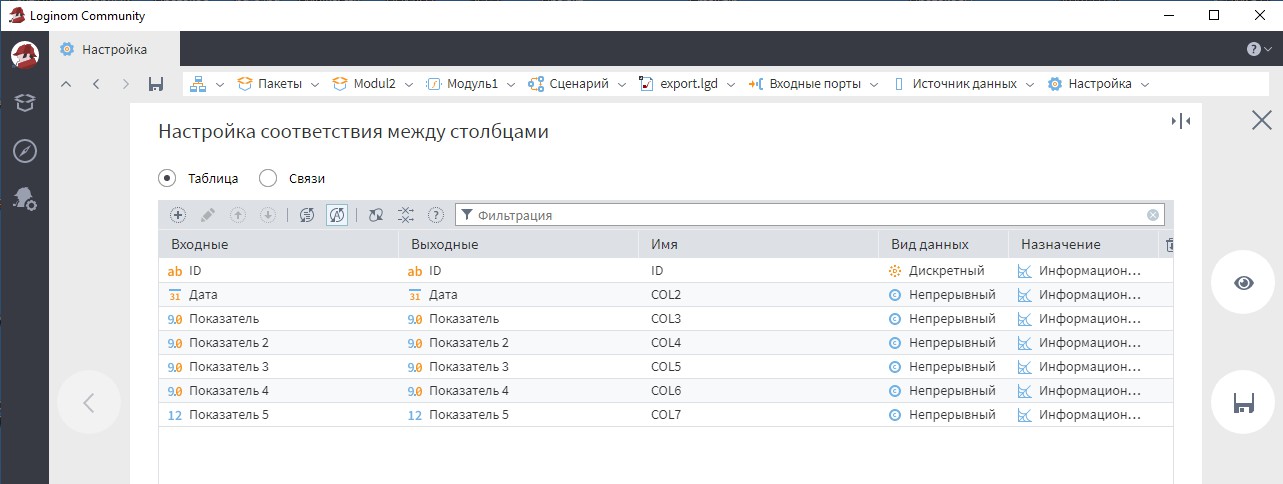
В результате получим простой сценарий, который импортирует файл в формате **txt** и экспортирует в формате **lgd**. На практике между узлами импорта и экспорта будет обработка.

Б. Посмотрим что получится, если изменится структура импортируемого файла. Создадим связь выходного порта файла **Несколько показателей** с входным портом узла экспорта **export.lgd**.

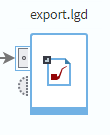
Мэппинг полей произошел автоматически так как была включена автосинхронизация входного порта. В том, что порт успешно сконфигурирован можно убедиться в его настройках.



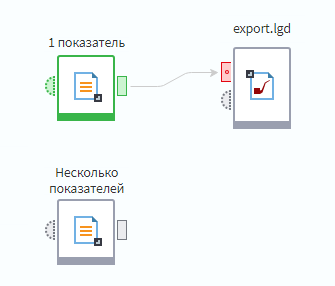
Зайдём в окно **Настройка соответствия между столбцами** узла **export.lgd**.



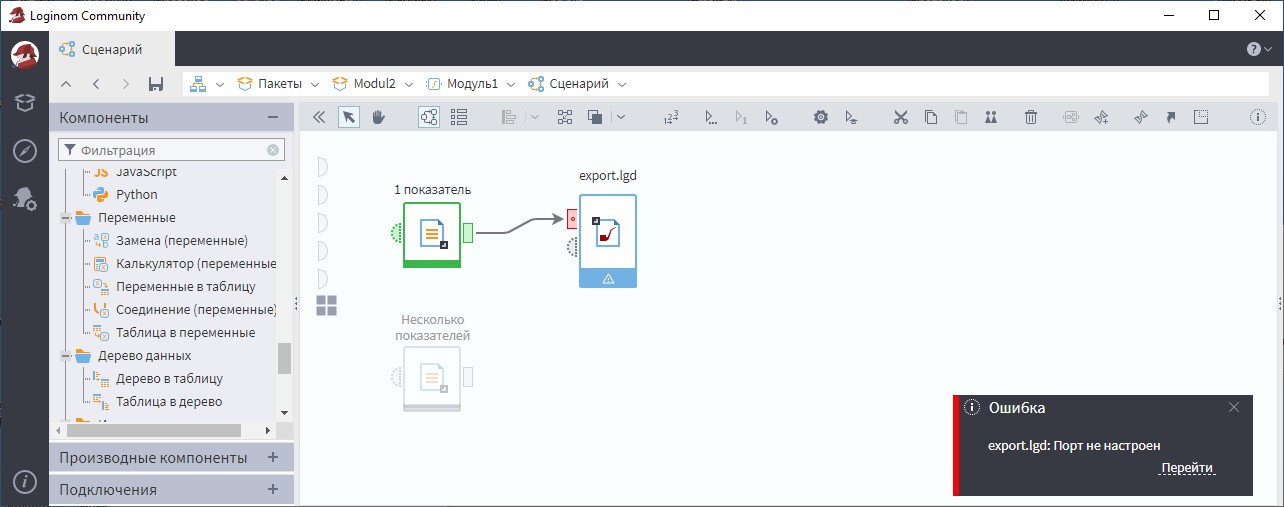
Можно заметить, что список входных полей порта обновился, расширился. Т.е. не пришлось ничего перенастраивать, изменяя структуру входного набора данных. Помогла автосинхронизация полей.

Выключим автоматическую синхронизацию в настройках входного порта узла **export.lgd** и сохраним изменения. На порте узла появится значок, сигнализирующий, что автосинхронизация выключена.

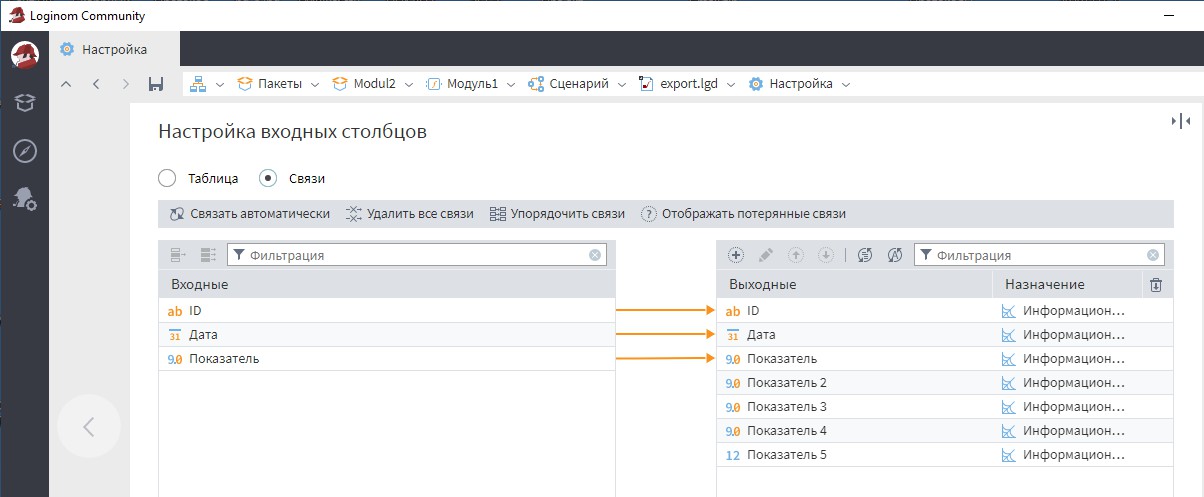
Вернём связь с набором данных **1 показатель**. Порт стал красного цвета. Он не сконфигурирован. Это произошло потому, что отключена автосинхронизация.



Если активировать узел, появится ошибка **Порт не настроен**.

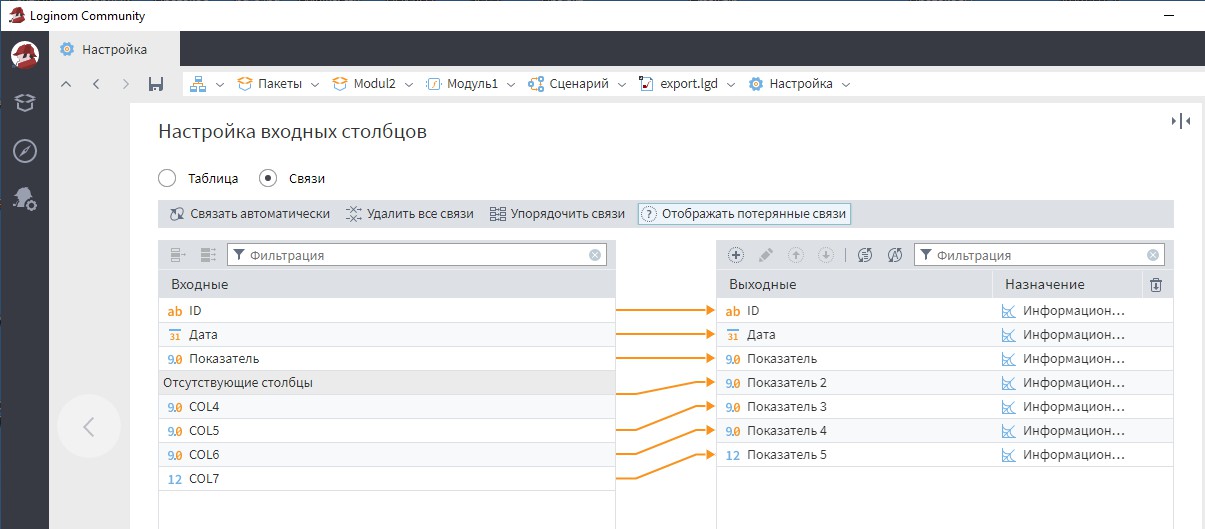


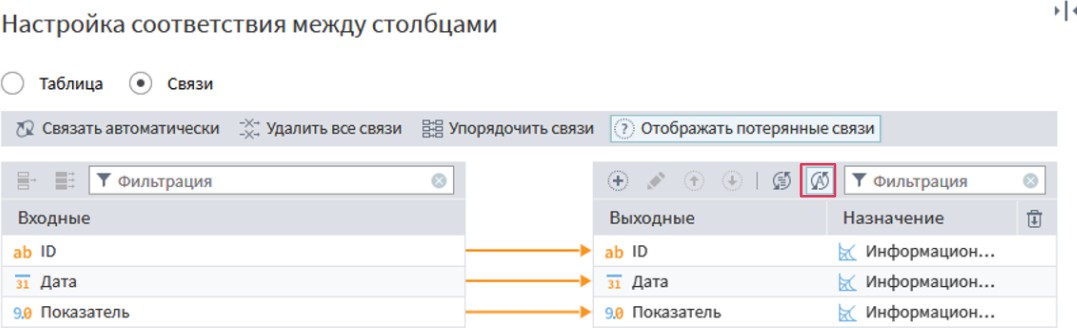
Зайдём в настройку порта узла экспорта. Посмотрим вкладку **Связи**.



На выходе стало 4 поля данных из старого набора данных (**Несколько показателей**) без связи. Это **потерянные связи**.

Чтобы отобразить потерянные связи, необходимо нажать на кнопку **Отображать потерянные связи** на панели инструментов.

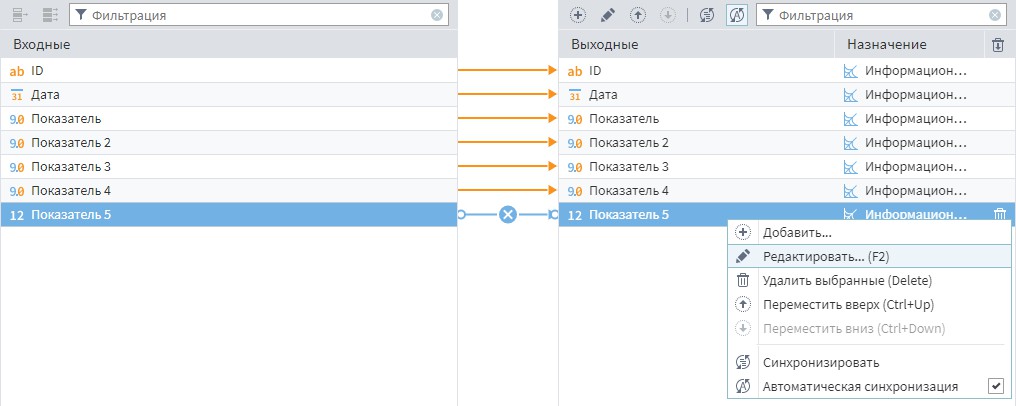


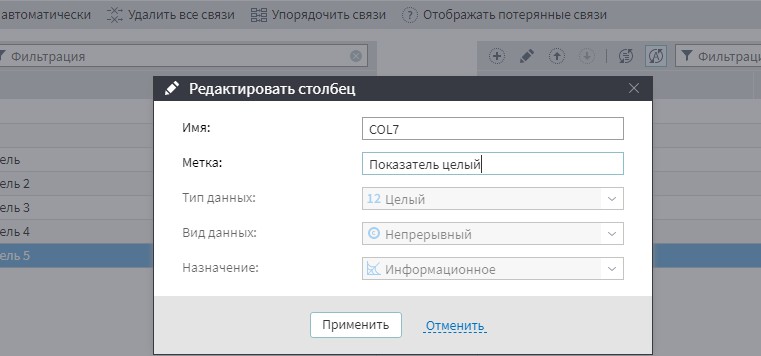
Мы работаем с документом **1 показатель**, поэтому необходимо удалить лишние поля, чтобы мэппинг прошел корректно и порт был настроен правильно.

Снова включим автосинхронизацию и сохраним настройки.

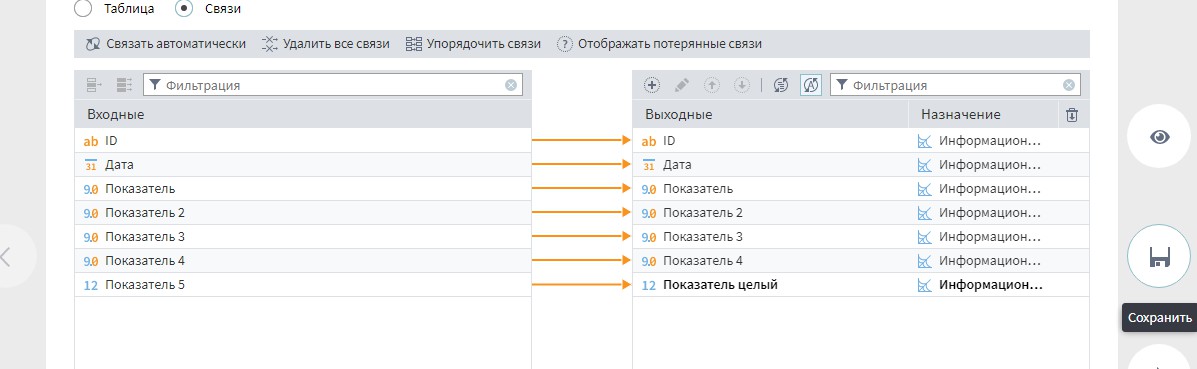
Поменяем связь. Соединим узел экспорта с набором **Несколько показателей**. Зайдём в настройку входного порта узла экспорта, зайдём во вкладку **Связи**. Откроем окно настройки столбца **Показатель 5** (**Редактировать столбец** из контекстного меню).

Во время мэппинга полей можно изменить метку столбца и имя. Тип и вид столбца можно изменить в случае, когда выходное поле не связано с входным.

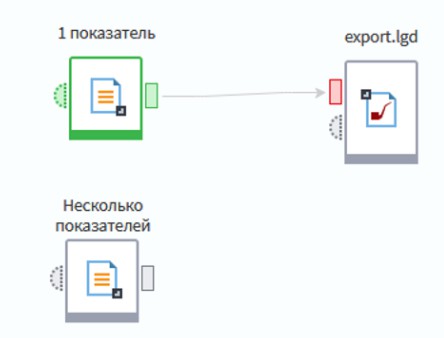


Изменим метку с **Показатель 5** на **Показатель целый**. Сохраним изменения, нажав на **Применить**.

Сохраним изменения и закроем мастер.



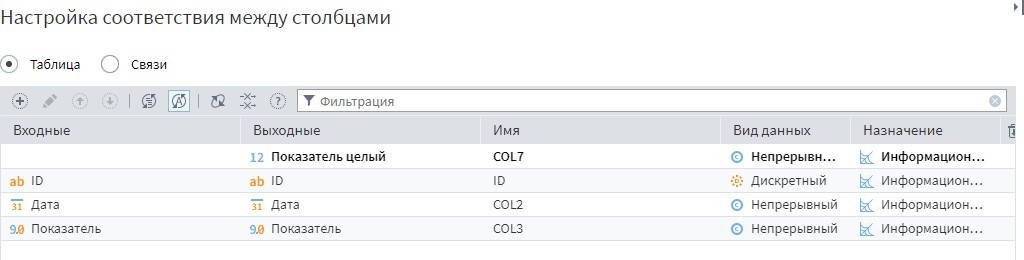
Вернём связь к узлу экспорту от набора данных **1 показатель**.



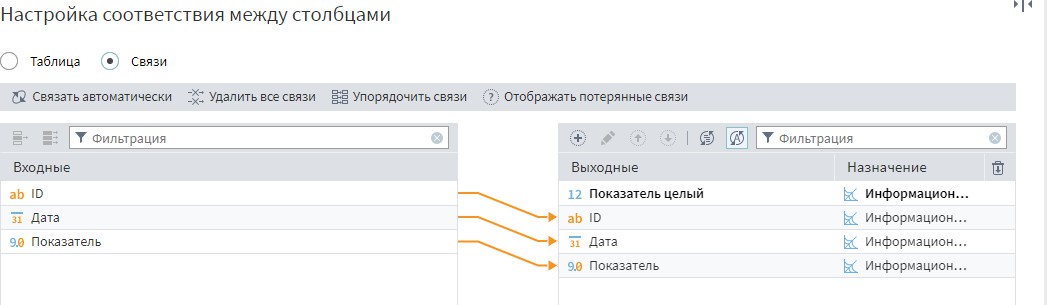
Даже при включенной автосинхронизации порт не был сконфигурирован.

Причины можно посмотреть в настройке порта узла экспорта.

Столбец **COL7** не синхронизировался.



Строка выделена жирным шрифтом.



Каждый столбец набора данных имеет признак: **обязательный**/ **необязательный**.

**Обязательные столбцы** – все столбцы, которые добавлены или изменены пользователем вручную.

Обязательные выходные столбцы при операции синхронизации в мэппинге не удаляются из списка. Если совпадут признаки **Имя+Тип**, такие столбцы будут связываться с входными. Разрешить конфликт мэппинга в случае, когда для обязательного поля не была найдена ни одна связь можно или удалив столбец из списка выходных, или создать связь вручную.

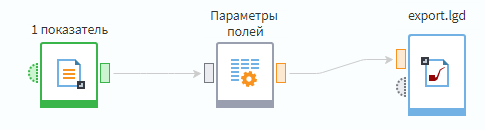
После изменения метки столбца на **Показатель целый**, столбец стал

# обязательным.

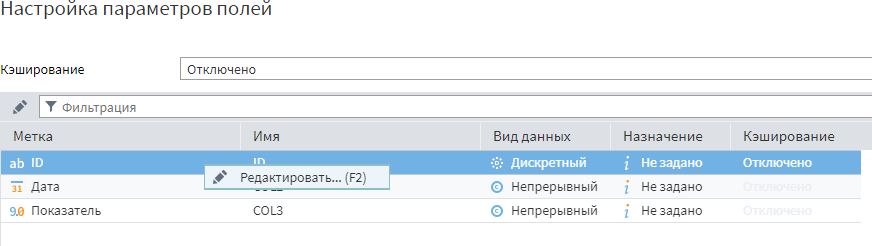
Чтобы разрешить конфликт, удалим столбец. После удаления любого столбца автосинхронизация отключается. Включим её.

Не всегда удобно изменять метаданные столбцов (имя, метку, тип и т.д.), удалять столбцы при мэппинге, потому что это меняет поведение автосинхронизации. В группе компонентов **Трансформация** есть специальный компонент – **Параметры полей**, предназначенный для этого.

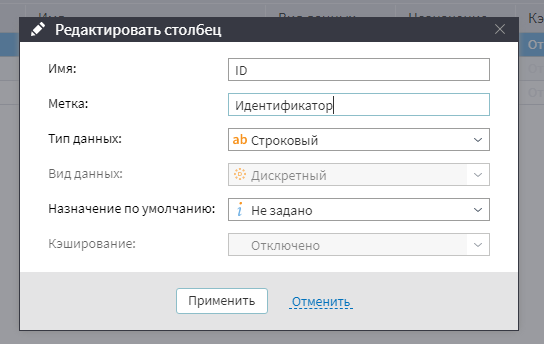
Поместим между узлами импорта и экспорта узел **Параметры полей**.

Входной порт узла **Параметры полей** благодаря автосинхронизации сконфигурировался, выходной порт частично сконфигурировался из-за того, что данные на него ещё не были поданы (узел будет сконфигурирован при активации узла).

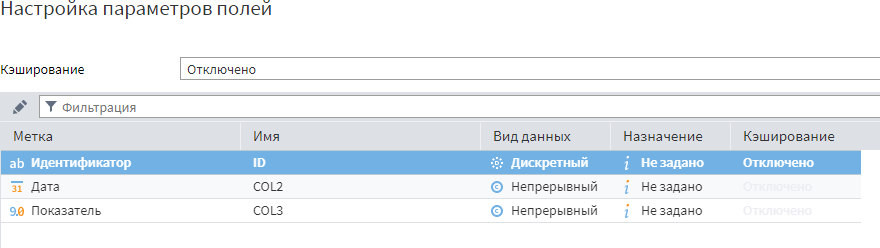
Зайдём в настройку узла **Параметры полей** и откроем окно настройки поля **ID**.



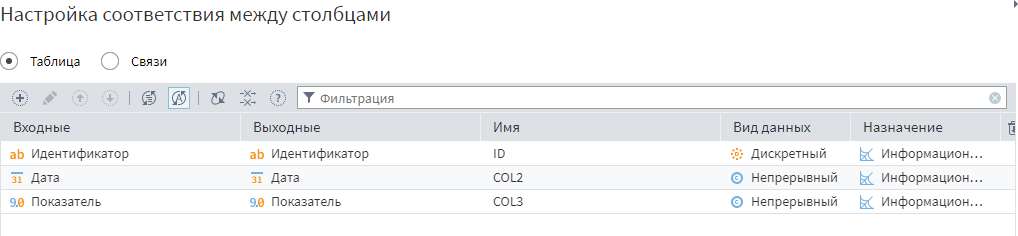
В окне **Редактировать столбец** заменим метку с **ID** на **Идентификатор**. Нажмём на **Применить**.



Убедимся, что столбец переименовался. Сохраним настройки. Выполним узел экспорта.



Посмотрим входной порт узла экспорта.



Столбец переименовался.

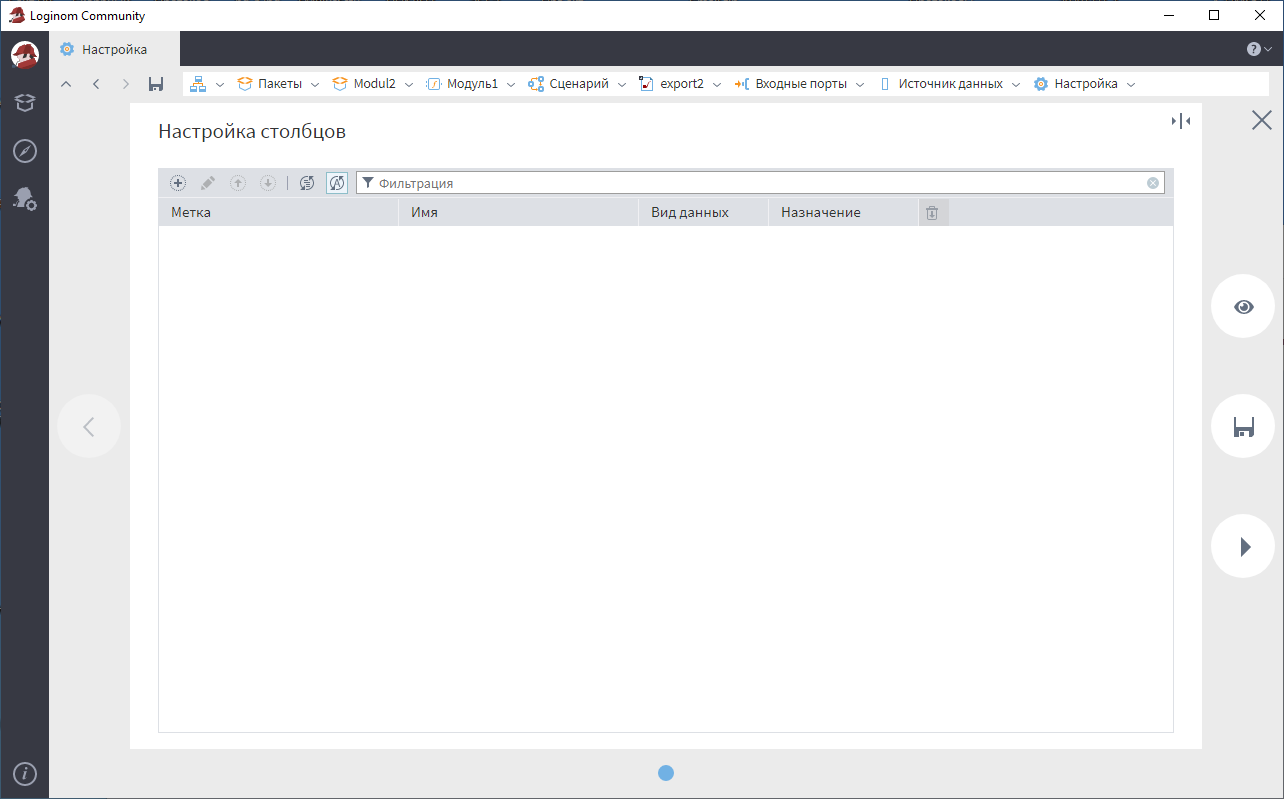
# Проектирование без данных

**Проектирование без данных** – это когда пользователь сам создаёт структуру набора данных во входном порте или портах, определяя метаданные столбцов.

# Практическое задание №10

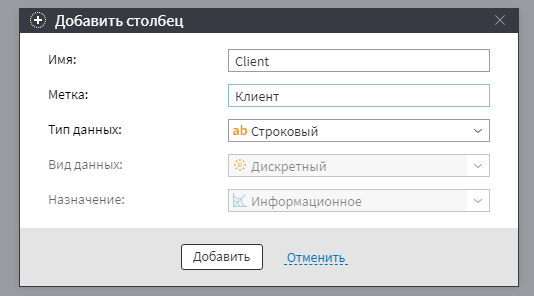
Добавим в область построения сценария узел экспорта **Loginom Data файл**.

Назовём его **export2.lgd**. Откроем настройку входящего порта. Окно настройки пустое.

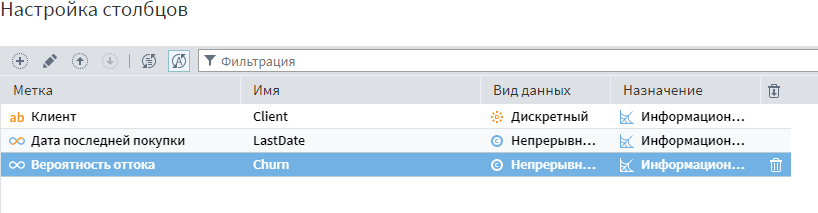


Нажмём кнопку добавить  на панели управления. Поменяем **имя** на **Client**,

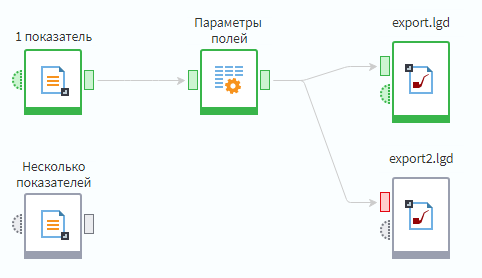
**метку** на **Клиент** и нажмём кнопку **Добавить**.



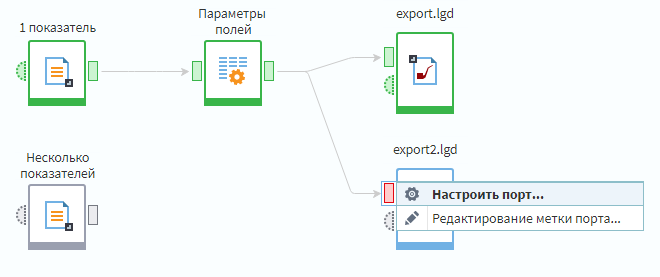
Таким же способом добавим столбцы **LastDate** и **Churn**.



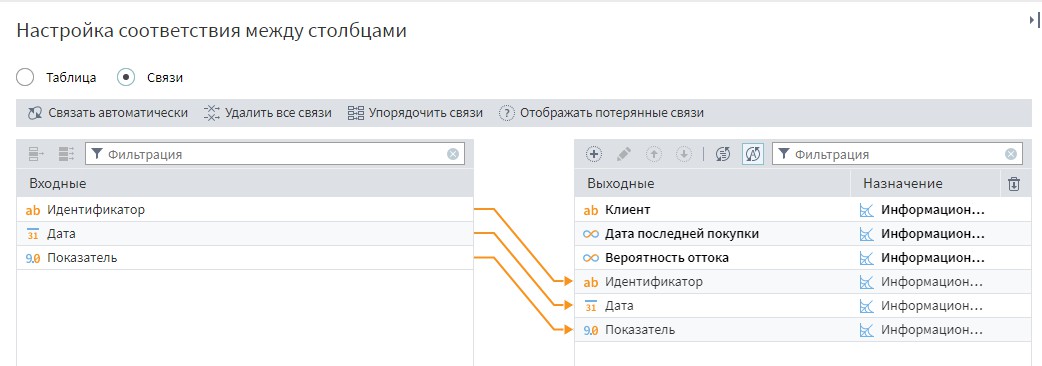
Соединим выходной порт узла **Параметры полей** и входной порт узла **eхport2.lgd**.

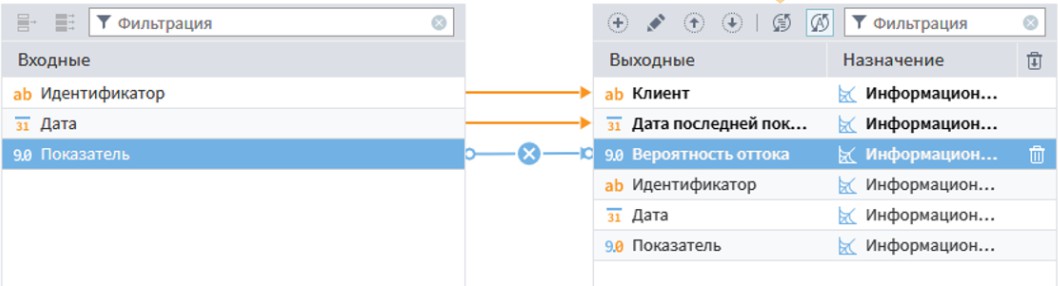


Откроем настройку порта.

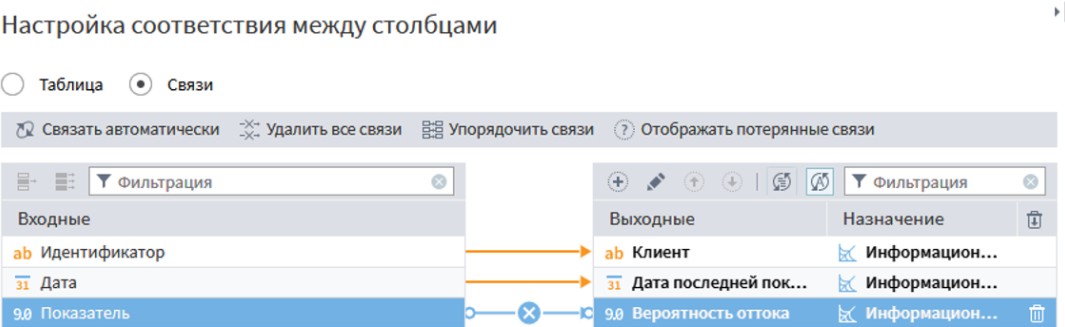


Сработала автосинхронизация: созданы новые столбцы.



Вручную поставим связи: Идентификатор -Клиент, Дата-Дата последней покупки, Показатель – Вероятность оттока.

Выполним синхронизацию, нажав на кнопку  .



Сохраним изменения.

**Итоги.** Был создан узел экспорта: задана структура набора данных, настроен мэппинг с имеющимися полями из другого узла. Метаданные столбцов (имена, метки) двух узлов различные. Если отключить автосинхронизацию в узле экспорта **export2.lgd**, то при подаче на его вход любых столбцов, имена и типы которых отличны от заданных нами, порт не сконфигурируется, узел при попытке активации выдаст ошибку.

Когда нужна автосинхронизация в узлах:

* Если нужно «протягивать» через узлы новые столбцы, которые могут появиться в структуре импортируемого набора данных, не выключайте автосинхронизацию.
* Для компонентов, которые сильно меняют структуру набора данных (Кросс-

таблица, Группировка), автосинхронизация не влияет на работу узлов.

* Автосинхронизация в выходном порте редко нуждается в отключении.
* Если есть необходимость строгого контроля какие должны быть поля на входах и на выходах узлов, не допускать изменения структуры обрабатываемых наборов данных, автосинхронизацию нужно отключить. При внесении изменений в сценарий и правке узлов может потребоваться больше усилий, направленных на ручную синхронизацию полей.