



Импорт из текстового файла

Используется для импорта данных из текстовых файлов с разделителями. Текстовый файл с разделителями представляет собой обычный текстовый файл, столбцы данных в котором разделены однотипными символами-разделителями, например, символами табуляции, запятыми, точками с запятой и т. д.

Вход

-  **Управляющие переменные (необязательный порт)** — переменными можно задать значения параметров мастера настройки.

Выход

-  **Набор данных (таблица данных).**

Мастер настройки

Настройка компонента происходит в три этапа:

Шаг 1. Импорт из текстового файла

- **Имя файла/URL** — путь к файлу, содержащему необходимые данные, или URL-ссылка для получения файла (поддерживаются протоколы http/https, а также HTTP Basic Authentication `https://login:password@hostname.com/path/file.csv`, `http://hostname.com/path/file.csv`).

Кнопка  вызывает диалог выбора файла.

Кнопка  позволяет добавить несколько файлов для импорта. Данные из всех файлов будут импортированы в одну итоговую таблицу. Файлы должны быть одинаковой структуры.

Для разделения файлов в строке используется специальный символ разделитель , например:

Имя файла 1.txt|Имя файла 2.txt

Имя файла может содержать маску, которая позволяет системе выполнять поиск файлов, соответствующих ей. Для маски применяется специальный символ . Например, `Путь до файла/М*ска.txt` или `Путь до файла/*.txt`. При задании маски нельзя использовать символ

`?`, т. к. он может являться частью URL-ссылки. Чтобы сохранить возможность множественного импорта с использованием URL-ссылок, данный символ не воспринимается как маска.

- **Параллельная обработка** — флаг, включающий или выключающий механизм параллельного импорта. При включении загрузка нескольких файлов происходит параллельно с использованием нескольких потоков, что повышает скорость множественного импорта.
- **Макс. кол-во потоков** — выбор количества потоков для параллельной обработки. Можно ввести значение самостоятельно или выбрать из предопределённых значений. Автоматический режим определяет оптимальное для каждого случая количество потоков.
- **Информация о файле** — выпадающий список, добавляет дополнительные поля, содержащие информацию о файле:
 - Нет — без дополнительных полей.
 - Имя файла — дополнительное поле с именем файла: `Товары.txt`.
 - Относительный путь — дополнительное поле содержит относительный путь до файла: `Пользователь/Downloads/Товары.txt`.
 - Абсолютный путь — дополнительное поле содержит абсолютный путь до файла: `C:\Users\Пользователь\Downloads\Товары.txt`.
 - Имя файла, время модификации — дополнительные поля, содержащие имя файла (поле Строкового типа) и время модификации файла (поле типа Дата/Время).
 - Относительный путь, время модификации — дополнительные поля, содержащие относительный путь (поле Строкового типа) и время модификации файла (поле типа Дата/время).
 - Абсолютный путь, время модификации — дополнительные поля, содержащие абсолютный путь (поле Строкового типа) и время модификации файла (поле типа Дата/время).
- **Кодовая страница** — кодировка, в которой хранится текст в файле.
- **Заголовок в первой строке** — флаг, определяющий, что метки полей создаваемого набора данных содержатся в первой строке импортируемого текста.
- **Пропустить строк** — количество строк с начала текста, не подлежащее импорту.

Шаг 2. Настройки форматов импорта

Определить автоматически — автоматически определяются все разделители и рассчитывается структура полей.

На этапе настройки указываются следующие параметры:

- **Разделитель столбцов** — символ, разделяющий данные соседних столбцов. Есть несколько стандартных разделителей:
 - Символ табуляции
 - Точка с запятой
 - Пробел
 - Запятая

Примечание: Если есть `tab` или `;`, используется тот, который чаще встречается, иначе приоритет отдаётся `,`. Также присутствует возможность задать разделитель вручную, для этого необходимо в строке заменить предложенный в форме ввода текст на нужный разделитель. Если названных разделителей нет, то делается попытка разделить по пробелам.

- **Считать последовательные разделители одним** — установка флага позволит принимать несколько подряд идущих разделителей столбцов за один.
- **Ограничитель строк** — символ, обрамляющий значения ячейки загружаемого набора данных. Текст, следующий за этим символом и предшествующий следующему его вхождению, импортируется как одно значение, даже если текст содержит символ-разделитель. В качестве ограничителя строк могут быть использованы следующие значения:
 - Нет
 - Двойная кавычка (`"`)
 - Одинарная кавычка (`'`)
 - Обратная кавычка (```)По умолчанию используется Двойная кавычка (`"`).
- **Пусто** — текст, представляющий значение Null. По умолчанию используется символ знака вопроса `?`. Доступные значения: *Пустая строка, ?, null, NULL*.
- **Переменный тип** — установка флага определяет поля, содержащие в себе записи с различными типами данных, как Переменный тип. При неактивном флаге таким полям присваивается Строковый тип.
- **Десятичный разделитель** — символ, разделяющий дробную и целую части числовых значений. Может принимать следующие значения:
 - Не задано
 - Точка (`.`)
 - Запятая (`,`)
- **Формат даты** — параметром задается формат даты, используемый в импортируемом тексте:
 - Не задано
 - `dd/mm/yyyy`
 - `mm/dd/yyyy`
 - `yyyy/mm/dd`
 - `dd/mm/yy`
 - `mm/dd/yy`
 - `yy/mm/dd`Во всех форматах: `dd` — день месяца в виде 2-х цифр, `mm` — номер месяца в виде 2-х цифр, `yy` — год в виде 2-х цифр, `yyyy` - год в виде 4-х цифр.
- **Разделитель даты** — символ, разделяющий компоненты даты. Может принимать следующие значения:
 - Не задано
 - Точка (`.`)

- Слэш (/)
- Обратный слэш (\)
- Дефис (-)
- **Разделитель времени** — символ, разделяющий компоненты времени. Может принимать следующие значения:
 - Не задано
 - Двоеточие (:)
 - Точка (.)

В параметрах *Десятичный разделитель*, *Разделитель даты* и *Разделитель времени* в качестве значений по умолчанию (Не задано) используются локальные настройки.

Примечание: Распознавание чисел и даты происходит так: выбирается первый строковый столбец, который преобразуется в нужный тип (целый, вещественный, дата/время). Разделитель, распознанный в этом столбце, будет применяться в остальной таблице.

- **Истина** — текст, представляющий логическое значение ИСТИНА. По умолчанию используется строка *True*. Доступные значения: *True*, *Истина*, *Да*.
- **Ложь** — текст, представляющий логическое значение ЛОЖЬ. По умолчанию используется строка *False*. Доступные значения: *False*, *Ложь*, *Нет*.
- **Автоопределение столбцов** — параметр, управляющий автоматической настройкой имён и типов данных полей при импорте. В выпадающем списке доступны 3 режима:
 - Нет — при импорте будут использоваться только те поля, которые были определены на этапе настройки мастера. Их структура строго фиксирована.
 - Только новые — если при импорте обнаружатся новые поля, которых не было на этапе настройки мастера, они будут прочитаны и автоматически настроены.
 - Все — поля настраиваются автоматически при каждом импорте данных. Табличная часть окна мастера и настройка соответствия между столбцами становятся недоступны для редактирования.

Табличная часть окна мастера предназначена для настройки параметров импорта для каждого поля набора данных.

-  **Обновить все** — рассчитывает все имена, метки, типы и виды данных полей.
-  **Определить типы данных** — рассчитывает заново только типы и виды данных.
- **Кол-во строк для анализа** — задает количество строк набора, по которым будет проводиться анализ типа данных. При этом максимальное число строк для отображения — 200.
- **Исходные данные** — в табличной части окна мастера будет показан исходный (текстовый) вариант загружаемых данных.
- **Результат** — в табличной части окна мастера будет показан результат преобразования данных с учетом установленных параметров полей.
-  **Развернуть** — разворачивает табличную часть мастера настроек.
-  **Свернуть** — сворачивает табличную часть мастера настроек.

Следующие настройки применяются для формирования результирующего набора данных вручную:

- **Имя** — задается уникальное имя поля.
- **Метка** — задает пользовательское описание поля.
- **Тип данных** — тип данных, в который будет преобразовано текстовое значение поля.
- **Вид данных** — вид данных, задается списком значений:
 - Непрерывный
 - Дискретный
 - Неопределенный
- **Использовать** — флаг включения поля в результирующий набор данных.

Шаг 3. Описание узла

На данном этапе мастера настройки узла доступен выбор режима формирования метки узла:

Автоматическая метка — метка генерируется системой на основе имени(имён) файла(ов) или управляющей переменной.

Для узла автоматическая метка формируется по следующим правилам:

- Если указано одно имя файла или одна маска, в метку помещается имя файла (или маска) без пути. Например, `Продажи.txt` .
- Если указано несколько файлов или масок, в метку помещается имя (или маска) первого файла и в скобках указывается общее количество элементов. Например, `Продажи_январь.txt (+2)` .
- Если имя файла задано через управляющую переменную, в метку подставляется имя этой переменной. Например, `%SalesFebruary%` .

Автоматически сформированная метка обновляется при изменении настроек импорта и отображается в сценарии без возможности редактирования напрямую. При необходимости можно принудительно обновить её через контекстное меню метки узла.

Пользовательская метка задаётся вручную и не меняется автоматически. В этом режиме можно ввести любое текстовое описание. Рядом с полем ввода отображается значок обновления, который становится активным, если текущая пользовательская метка отличается от метки компонента по умолчанию. Нажатие на значок обновления восстанавливает метку компонента по умолчанию.

Режим формирования метки можно изменить как в мастере настройки узла (через выпадающий список на третьем шаге мастера настройки), так и через контекстное меню метки узла в сценарии (правой кнопкой мыши по метке узла).

У скопированного или клонированного узла сохраняется режим формирования метки.

Если у исходного узла метка формируется автоматически, то у скопированных и клонированных узлов не будет добавляться индекс.

Автоматическое определение структуры таблицы

Свойства таблицы, которые могут быть определены автоматически:

- Заголовок в первой строке

- Ограничитель строк
- Десятичный разделитель
- Разделитель даты
- Разделитель времени
- Формат даты

Особенности множественного импорта файлов

При множественном импорте настройки мастера являются общими для всех файлов, которые будут импортироваться. Активация узла будет читать строки из всего диапазона файлов, пока не будет прочитан последний.

Чтение имени файла выполняется либо до разделителя, либо до конца строки. Из этого следует, что если в конце строки *Имя файла/URL* оставить разделитель, то программа выполнит пустое чтение в конце обработки строки. В связи с этим пустое чтение не воспринимается как ошибка. Таким образом, ситуация `имя1| |имя2` и `имя3|` не считается ошибочной.

При множественном импорте на второй странице мастера настройки отображаются строки из нескольких файлов, пока их общий диапазон строк не превысит указанный лимит (по умолчанию для импорта из текстового файла это 25 строк).

При определении типов колонок в учёт берутся строки из нескольких файлов.

Статьи в разделе:

- [Многострочные значения](#)