

Департамент образования и науки города Москвы Государственное
автономное образовательное учреждение высшего образования города
Москвы «Московский городской педагогический университет» Институт
цифрового образования Департамент информатики, управления и технологий

ДИСЦИПЛИНА:

Инструменты для хранения и обработки больших данных

Практическая работа №3

Тема:

Архитектура хранилищ данных: традиционная и облачная

Выполнила: Шепелева Е. В., группа: АДЭУ-201

Преподаватель: Т. М. Босенко

Москва

2022

ETL Pentaho DI

№1 Mail

▼ Mail

- Get mails (POP3/IMAP)
- Mail**
- Mail validator

Mail

Name of mail job entry: Mail

Addresses | Server | EMail Message | Attached Files

Destination

Destination address:

Cc:

Bcc:

Sender

Sender name

Sender address:

Reply to

Contact person:

Contact phone:

Help OK Cancel

Используйте запись задания «Mail», чтобы отправить текстовое или HTML-сообщение электронной почты с дополнительными вложенными файлами. Эта запись задания используется в конце выполнения задания в большинстве случаев. Его можно использовать, чтобы объявить, как о неудаче, так и об успехе задания. Например, нередко в конце успешной загрузки отправляется электронное письмо в список рассылки с уведомлением об успешной загрузке и включением файла журнала. Если есть ошибки, можно отправить электронное письмо, чтобы предупредить людей из списка рассылки.

Подробное описание вкладок


Addresses tab


| Вариант | Определение |
|---------------------|--|
| Job entry name | Уникальное имя записи задания на холсте. Запись о задании может быть размещена на холсте несколько раз; однако это будет та же запись о работе |
| Destination Address | Место назначения для электронной почты; вы можете указать несколько адресов, если разделите их пробелом. |
| Сс: | Идентичная копия сообщения также отправляется на все адреса, указанные в поле Копия:. Чтобы ввести более одного адреса в поле Копия:, разделите их пробелом. |
| ВСс: | Отправить получателю, чей адрес электронной почты не указан в сообщении |
| Sender name | Имя человека, отправляющего электронное письмо |
| Sender address | Адрес электронной почты лица, отправляющего электронное письмо |
| Reply to | Электронный адрес человека, которому отправить ответ |
| Contact person | Имя контактного лица, которое будет указано в письме |
| Contact phone | Контактный телефон для размещения в письме |


Server tab


| Вариант | Определение |
|----------------------------|---|
| SMTP Server | Адрес SMTP-сервера |
| Port | Порт, на котором работает SMTP-сервер. |
| Use Authentication | Разрешить использовать аутентификацию на SMTP-сервере |
| Authentication user | Имя учетной записи пользователя SMTP |
| Authentication password | Пароль учетной записи пользователя SMTP |
| Use Secure Authentication? | Включить безопасную аутентификацию |
| Secure Connection Type | Выберите тип аутентификации (SSL, TLS) |


№2 Copy files

▼  File management

 Add filenames to result

 Compare folders

 Convert file between Windows and Unix


 Copy files

Copy files

— □ ×

Job entry name:

Copy files



Files Settings

Files/Folders:

×

| # | Source Environment | Source File/Folder | Wildcard (RegExp) | Destination Environment | Destination File/Folder |
|---|--------------------|--------------------|-------------------|-------------------------|-------------------------|
| 1 | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

Help

OK Cancel

Вы можете скопировать один или несколько файлов, или папок с этой записью задания.

Подробное описание вкладок Files

| Вариант | Определение |
|-----------------------|---|
| Job Entry Name | Указывает имя шага. |
| Source Environment | Указывает файловую систему или конкретный кластер, в котором можно найти элемент, который вы хотите ввести. Возможные варианты: Локальный и <Статический>. Local: указывает, что элемент, указанный в поле File/Folder, находится в файловой системе, которая является локальной для Spoon. <Static>: указывает, что элемент, указанный в поле File/Folder, должен точно использовать имя пути в этом поле. Используйте это, если вы уже знаете путь к файлу и хотите просто скопировать и вставить его в окно. |
| Source File/Folder* * | Файл или каталог для копирования. Если вы выберете вариант, отличный от <Статический> в |

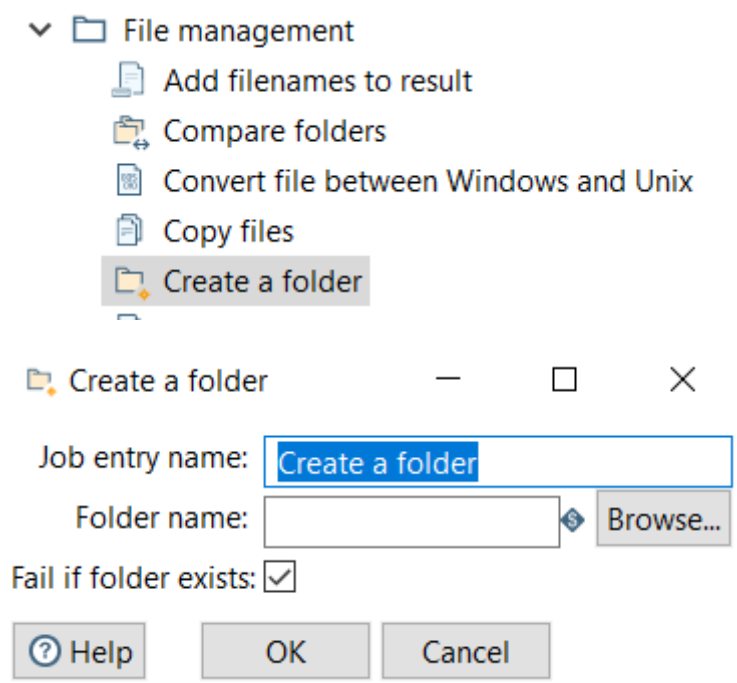
| | |
|-------------------------|---|
| | поле Исходная среда , нажмите кнопку в поле, чтобы открыть окно Открыть файл . |
| Wildcard | Определяет файлы, которые копируются в терминах регулярных выражений (вместо статических имен файлов), например: *.*\txt будет любым файлом с расширением .txt. |
| Destination Environment | <p>Указывает файловую систему или конкретный кластер, в который вы хотите поместить файл. Возможные варианты: Локальный и <Статический>.</p> <p>Local: указывает, что элемент, указанный в поле File/Folder, находится в файловой системе, которая является локальной для Spoon.</p> <p><Static>: указывает, что элемент, указанный в поле File/Folder, должен точно использовать имя пути в этом поле. Используйте это, если вы уже знаете путь к файлу и хотите просто скопировать и вставить его в окно.</p> |
| Destination File | Указывает имя целевой среды. Если в поле «Среда назначения» выбран вариант, отличный от «Статический» , нажмите кнопку в поле, чтобы открыть окно «Открыть файл» . |

Settings

| | |
|---------------------------|--|
| | |
| Include Subfolders | Если выбрано, все подкаталоги в выбранном каталоге также будут скопированы. |
| Destination is a file | Определяет, является ли пункт назначения файлом или каталогом |
| Copy empty folders | Если выбрано, будут скопированы все каталоги, даже если они пусты. Для того, чтобы этот параметр был действительным, необходимо выбрать параметр « Включить подпапки» . (Подстановочный знак ДОЛЖЕН быть пустым, чтобы это работало) |
| Create destination folder | Если выбрано, будет создан указанный целевой каталог, если он в настоящее время не существует. |

| | |
|------------------------------------|---|
| Replace existing files | Если выбрано, дубликаты файлов в целевом каталоге будут перезаписаны. |
| Remove source files | Если выбрано, удаляет исходные файлы после копирования (процедура перемещения). |
| Copy previous results to arguments | Копирует предыдущие результаты в аргументы. |
| Add files to result files name | Любые скопированные файлы появятся в результате этого шага; показывает список файлов, которые были скопированы на этом шаге |

№3 Create a folder
















Эта запись задания просто создает папку в указанном месте.


Подробное описание вкладок

| Вариант | Определение |
|----------------|--|
| Job entry name | Имя записи о работе. Примечание. Это имя должно быть уникальным в рамках одного задания. Запись задания может быть размещена на холсте несколько раз, однако это будет одна и та же запись задания. |
| Folder name | Имя создаваемой папки (может быть полный путь) |

| | |
|-----------------------|---|
| Fail if folder exists | Если вы включите этот параметр, эта запись задания завершится ошибкой, если указанная папка уже существует. |
|-----------------------|---|

№4 HTTP

- ▼  File management
 -  Add filenames to result
 -  Compare folders
 -  Convert file between Windows and Unix
 -  Copy files
 -  Create a folder
 -  Create file
 -  Delete file
 -  Delete filenames from result
 -  Delete files
 -  Delete folders
 -  File compare
 -  HTTP

 HTTP
 —
□
×

Name of job entry:

General

Headers

URL:

Run for every result row? ☐

Input field which contains URL

Input field which contains upload

Input field which contains

Authentication

Username:

Password:

Proxy server for upload:

Proxy port:

Ignore proxy for hosts:

Upload file

Upload file:

Webservice reply


Target file:

Append to specified target file? ☐

Add date and time to file name? ☐

Target file extension:

Add filename to result filenames ☒

 Help

OK

Cancel

Используйте запись задания HTTP для получения файла с веб-сервера по протоколу HTTP. Эта запись задания может использоваться для доступа к данным на партнерских веб-сайтах. Например, ежедневный экспорт данных или ежедневный список клиентов находится на указанном веб-сайте. Кроме того, поставщики SaaS могут предоставить вам URL-адрес для поиска отчета. Вы можете вызвать этот URL-адрес, чтобы получить файл Excel или ZIP-файл, содержащий данные. Salesforce требует, чтобы вы использовали API-интерфейсы SOAP для извлечения данных.

















Если HTTP-трафик в вашей корпоративной среде слишком велик, вы можете использовать прокси-сервер с HTTP-аутентификацией.

Подробное описание вкладок

| Вариант | Определение |
|--|---|
| Name of job entry | Имя записи о работе. Это имя должно быть уникальным в рамках одного задания. Запись задания может быть размещена на холсте несколько раз, однако это будет одна и та же запись задания. |
| URL | URL-адрес HTTP файла, который необходимо получить, или имя каталога для хранения загруженного файла |
| Run for every result row? | Установите этот флажок, если вы хотите запускать эту запись задания для каждой строки, созданной предыдущим преобразованием. Воспользуйтесь функцией «Копировать строки в результат». Если выбрано, HTTP-запрос будет сделан для каждого результата. В противном случае файл извлекается только один раз. |
| Input field which contains URL | Если «Выполнить для каждой строки результата?» выбрана опция, указанное здесь поле будет определять URL-адрес файла для каждой строки |
| Input field which contains upload file name | Если «Выполнить для каждой строки результата?» выбран вариант, указанное здесь поле будет определять локальный файл, который будет загружен на URL-адрес, связанный с полем «Поле ввода, которое содержит URL-адрес». |
| Input field which contains destination file name | Если «Выполнить для каждой строки результата?» выбран вариант, указанное здесь поле будет определять локальный файл, в который загружается результат с URL-адреса, связанного с полем «Поле ввода, которое содержит URL-адрес» |
| Username | Если сайт требует аутентификации, используйте это имя пользователя для входа |
| Password | Если имя пользователя определено, это пароль для него |
| Proxy server for upload | URL-адрес прокси-сервера, который вы хотите подключить к URL-адресу HTTP через |
| Proxy port | Если прокси-сервер определен, это номер порта, который он прослушивает. |
| Ignore proxy for hosts | Список исключений регулярного выражения для перенаправления прокси. Это может быть полезно при работе в интрасети или VPN. |
| Upload file | Если вы загружаете файл, это будет его имя на удаленном сервере. |
| Target file | Если вы загружаете файл, это его имя в вашей локальной файловой системе |

| | |
|----------------------------------|---|
| Append specified file? to target | Если выбрано и целевой файл уже существует, PDI добавит все новые данные в конец файла. |
| Add date and time to file name? | Если выбрано, дата и время HTTP-запроса (в формате ггггММдд_ЧЧммсс) будут добавлены к имени целевого файла. |
| Target file extension | Если выбран предыдущий вариант, в этом поле указывается расширение (буквы после точки) целевого имени файла. |
| Add filename to result filename | Любые скопированные файлы появятся в результате этого шага; показывает список файлов, которые были скопированы на этом шаге |

№5 Unzip file

- ▼  File management
 -  Add filenames to result
 -  Compare folders
 -  Convert file between Windows and Unix
 -  Copy files
 -  Create a folder
 -  Create file
 -  Delete file
 -  Delete filenames from result
 -  Delete files
 -  Delete folders
 -  File compare
 -  HTTP
 -  Move files
 -  Process result filenames
 -  Unzip file

Unzip file

— □ ×

Job entry name: Unzip file

General

Advanced

Source files

Get args from previous

Zip File name:

Files...

Folders...

Source Wildcard (RegExp)

Unzipped files

Use zipfile name as root

Target Directory:

Folders...

Create folder

Include Wildcard (RegExp):

Exclude Wildcard (RegExp):

Include date in filename

Include time in filename

Specify Date time format

Date time format

Add original timestamp

Set modification date to original

If file exists

Skip

After extraction

Do Nothing

Move Files To

Folders...

Create folder

Help

OK

Cancel

Вы можете разархивировать один или несколько файлов с помощью этой записи задания.


Подробное описание вкладок


| Вариант | Определение |
|----------------|--|
| Job entry name | Имя записи о работе. |
| Source files | В этом разделе вы можете указать список исходных (zip) файлов Получить аргументы из предыдущего: отметьте этот параметр, если хотите использовать список файлов результатов (из предыдущей записи задания) в качестве списка файлов для распаковки (см. ошибку PDI-9819 — вместо файлы результатов) |


| | |
|----------------|--|
| | <p>Имя Zip-файла: имя ZIP-файла или папки, если вы хотите использовать подстановочный знак.</p> <p>Исходный подстановочный знак: если предыдущий вариант — это папка, вы можете ввести здесь подстановочный знак регулярного выражения. -</p> <p>Примечание. Регулярное выражение сравнивается с абсолютным путем к файлу, и ДОЛЖНО быть выполнено полное совпадение.</p> <p>Например, для сопоставления "/folder/test.zip" регулярное выражение "te. \.zip" окажется пустым. Вместо этого используйте ".*te. \.zip" для учета папок перед ним.** Для путей к файлам Windows направление косой черты будет изменено на противоположное при совпадении, поэтому используйте "\\" вместо "\\". ".</p> |
| Unzipped files | <p>Этот раздел позволяет нам указать, что делать с разархивированными файлами.</p> <p>Использовать имя zip-файла в качестве корневого каталога: установите этот флажок, если хотите создать отдельный каталог для каждого имени zip-файла (то же имя, что и у файла).</p> <p>Целевой каталог: целевой каталог для распаковки.</p> <p>Создать папку: установите этот флажок, если хотите создать целевую папку</p> <p>Включить подстановочный знак : используйте это регулярное выражение, чтобы выбрать файлы в zip-архивах для извлечения.</p> <p>Exclude wildcard : используйте это регулярное выражение, чтобы выбрать файлы в zip-архивах для извлечения.</p> <p>Включить дату в имя файла: включить текущую дату в имена разархивированных файлов (в формате ггггММдд)</p> <p>Включить время в имя файла: Включить время (в формате ЧЧммсс)</p> <p>Укажите формат даты и времени самостоятельно: позволяет вам самостоятельно указать формат даты и времени (по умолчанию: ггггММдд'_'ЧЧммсс)</p> <p>Если файлы существуют: выберите действие, которое необходимо предпринять, если целевой (разархивированный) файл существует: пропустить, перезаписать и т. д.</p> <p>После извлечения: выберите действие после извлечения zip-файла: Ничего не делать, Удалить файлы, Переместить файлы.</p> <p>Переместить файлы в : если предыдущая опция «Переместить файлы», вы можете выбрать здесь целевой каталог.</p> |

| | |
|----------|---|
| Advanced | <p>Добавить извлеченный файл в результат: добавьте имена извлеченных файлов в список файлов результатов этой записи задания для использования в следующих записях задания.</p> <p>Условие успеха: позволяет указать фактор успеха этой записи задания: когда все прошло хорошо, или если было только X ошибок, или если было извлечено не менее Y файлов.</p> |
|----------|---|

№6 Check Db connections

▼  Conditions

 Check DB connections

 Check DB connections


Job entry name

Connections

| # | Connection | Wait | Units of time |
|---|------------|------|---------------|
| 1 | | | |

Get connections

Delete

 Help

OK


Cancel


Эта запись задания позволяет проверить подключение к одной или нескольким базам данных.


Подробное описание вкладок


| Вариант | Определение |
|-----------------|--|
| Job entry name | Имя этой записи в том виде, в котором оно отображается в рабочей области преобразования. |
| Connection | Список подключений. |
| Wait | После открытия соединения подождите x (с, мин, час). |
| Units of Time | Укажите единицу измерения времени, в течение которого необходимо оставаться на связи. |
| Get connections | Получите доступные подключения. |

№7 SQL

▼  Scripting

 JavaScript

 Shell

 SQL

SQL

Job entry name:

Connection

SQL from file ☐

SQL filename

Send SQL as single ☐

Use variable substitution? ☐

SQL Script:

Line 1 Column 0

Используйте запись задания SQL для выполнения сценария SQL. Вы можете выполнить несколько операторов SQL, если они разделены точкой с запятой. Запись задания SQL является гибкой; вы можете выполнять вызовы процедур, создавать и анализировать таблицы и многое другое. Общие области применения, связанные с записью задания SQL, включают усечение таблиц, удаление индекса, загрузку разделов, обновление материализованных представлений, отключение ограничений, отключение статистики и т. д.

Подробное описание вкладок

| Вариант | Определение |
|-------------------------------|---|
| Job entry name | Имя этой записи в том виде, в котором оно отображается в рабочей области преобразования. |
| Database Connection | Используемое соединение с базой данных. |
| SQL from file | Включите эту опцию, чтобы загрузить оператор SQL из файла, заданного именем файла SQL. |
| SQL filename | Имя файла с операторами SQL. |
| Send SQL as single statement? | Включите эту опцию, чтобы не разделять выражение точкой с запятой. Это часто полезно, когда задан сценарий или несколько операторов должны быть обработаны и зафиксированы как один оператор. |
| Use variable substitution? | Позволяет использовать переменные в сценарии SQL. |
| SQL script | SQL-скрипт для выполнения. Несколько операторов могут быть разделены точкой с запятой. |

№8 Hadoop copy files

▼ Big Data

- Amazon EMR job executor
- Amazon Hive job executor
- Hadoop copy files

Hadoop copy files

Job entry name:
Hadoop copy files

Files Settings

| # | Source Environment | Source File/Folder | Wildcard (RegExp) | Destination Environment | Destination File/Folder |
|---|--------------------|--------------------|-------------------|-------------------------|-------------------------|
| 1 | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

Help OK Cancel

Эта запись задания копирует файлы в кластере Hadoop из одного места в другое.

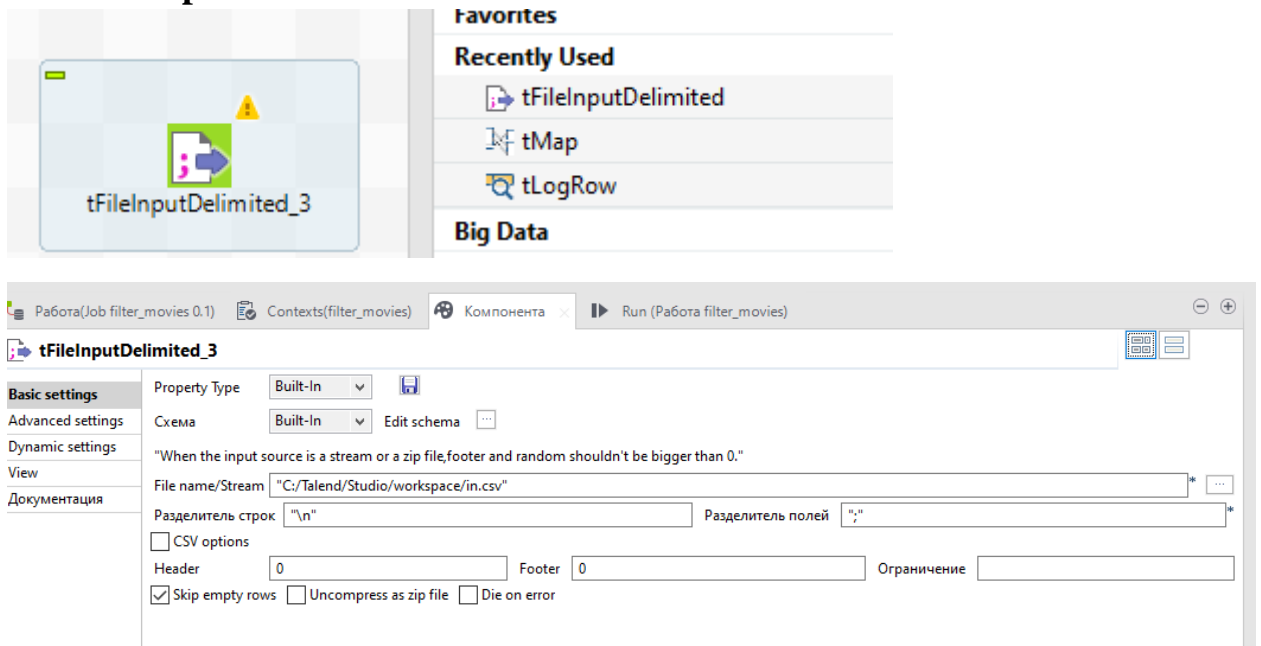
Подробное описание вкладок

| Вариант | Определение |
|-------------------------------|---|
| Include Subfolders | Если выбрано, все подкаталоги в выбранном каталоге также будут скопированы. |
| Destination is a file | Определяет, является ли пункт назначения файлом или каталогом |
| Copy empty folders | Если выбрано, будут скопированы все каталоги, даже если они пусты, чтобы этот параметр был действительным, необходимо выбрать параметр « Включить подпапки» . |
| Create destination folder | Если выбрано, будет создан указанный целевой каталог, если он в настоящее время не существует. |
| Replace existing files | Если выбрано, дубликаты файлов в целевом каталоге будут перезаписаны. |
| Remove source files | Если выбрано, удаляет исходные файлы после копирования (процедура перемещения). |
| Copy previous results to args | Если выбрано, будут использоваться результаты предыдущего шага в качестве источников и пунктов назначения. |
| File/folder source | Файл или каталог для копирования; нажмите «Обзор» и выберите «Hadoop» , чтобы ввести сведения о подключении к кластеру «Hadoop». |

| | |
|-------------------------|---|
| File/folder destination | Файл или каталог для копирования; нажмите «Обзор» и выберите «Hadoop» , чтобы ввести сведения о подключении к кластеру «Hadoop». |
| Wildcard (RegExp) | Определяет файлы, которые копируются в терминах регулярных выражений (вместо статических имен файлов), например: *.*\txt будет любым файлом с расширением .txt. |
| Files/folders | Список выбранных источников и мест назначения |

Talend

№1 tFileInputDelimited



Считывает файл с разделителями построчно, чтобы разбить его на поля, а затем отправляет поля, как определено в схеме, следующему компоненту.

Подробное описание вкладок

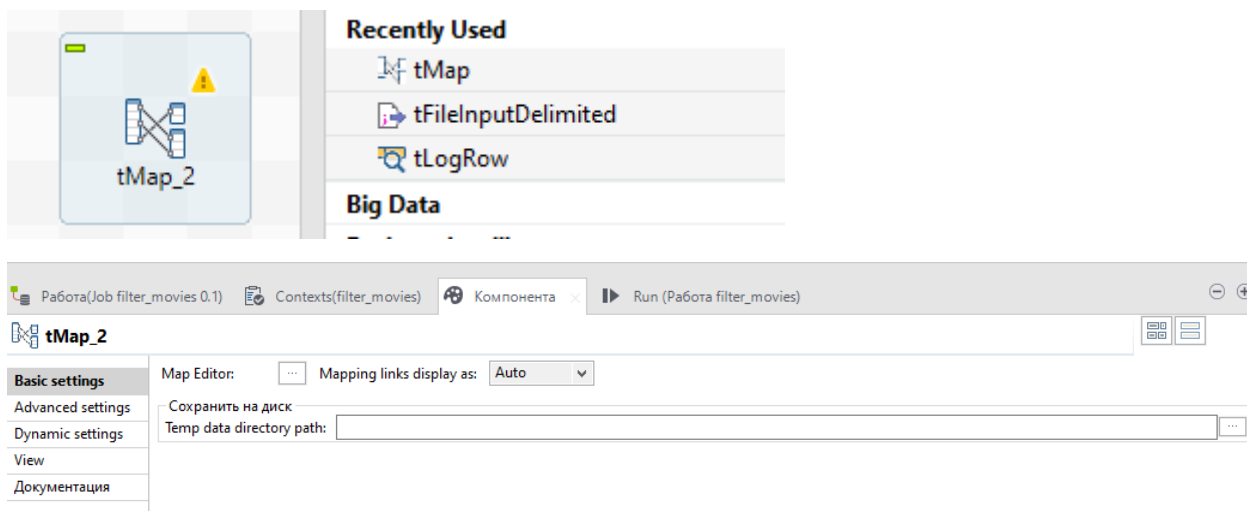
| Вариант | Определение |
|------------------|---|
| | Встроенная : данные о свойствах не хранятся централизованно. |
| | Репозиторий : выберите файл репозитория, в котором хранятся свойства. |
| File Name/Stream | Имя файла : имя и путь к файлу, который необходимо обработать. Поток: поток данных для обработки. Данные должны быть добавлены в поток, чтобы tFileInputDelimited мог получить эти данные через соответствующую репрезентативную переменную. |

| | |
|-----------------|---|
| | <p>Эта переменная может быть уже предопределена в вашей Studio или предоставлена контекстом или компонентами, которые вы используете вместе с этим компонентом; в противном случае вы можете определить его вручную и использовать в соответствии с дизайном вашего задания, например, с помощью tJava или tJavaFlex .</p> <p>Чтобы избежать неудобств рукописного ввода, вы можете выбрать интересующую переменную из списка автодополнения (Ctrl+Space), чтобы заполнить текущее поле при условии, что эта переменная была правильно определена.</p> |
| Row separator | Разделитель, используемый для обозначения конца строки. |
| Field separator | Введите символ, строку или регулярное выражение, чтобы разделить поля для передаваемых данных. |
| CSV options | <p>Установите этот флажок, чтобы указать следующие параметры CSV:</p> <p>Escape char : введите управляющий символ между двойными кавычками.</p> <p>Текстовое вложение : введите символ вложения (только один символ) между двойными кавычками. Например, "" необходимо вводить, когда в качестве символа включения используются двойные кавычки (").</p> <p>Рекомендуется использовать стандартный escape-символ, то есть "\". В противном случае вы должны установить один и тот же символ для Escape char и Text вложения . Например, если escape-символ установлен на "\", текстовое вложение может быть установлено на любой другой символ. С другой стороны, если escape-символ установлен на другой символ, а не на "\", текстовое вложение может быть установлено на любые другие символы. Однако escape-символ будет заменен на тот же символ, что и текстовое вложение. Например, если escape-символ установлен на "#", а текстовое вложение установлено на "@", escape-символ будет изменен на "@", не "#".</p> |
| Header | <p>Введите количество строк, которые необходимо пропустить в начале файла.</p> <p>Обратите внимание, что при использовании динамической схемы первая строка входных данных всегда обрабатывается как строка заголовка, независимо от того, установлено значение поля заголовка или нет. Для получения дополнительной информации о динамической схеме см. Руководство пользователя Talend Studio .</p> |

| | |
|------------------------|--|
| | |
| Footer | Количество строк, которые необходимо пропустить в конце файла. |
| Limit | Максимальное количество обрабатываемых строк. Если Limit = 0, ни одна строка не считывается и не обрабатывается. |
| Schema and Edit Schema | <p>Схема представляет собой описание строки. Он определяет количество полей (столбцов), которые необходимо обработать и передать следующему компоненту. При создании задания Spark избегайте зарезервированного слова line при именовании полей.</p> <p>Щелкните Изменить схему, чтобы внести изменения в схему. Если текущая схема имеет тип Репозиторий, доступны три варианта:</p> <p>Просмотреть схему: выберите этот параметр, чтобы просмотреть только схему.</p> <p>Изменить на встроенное свойство: выберите этот параметр, чтобы изменить схему на встроенную для локальных изменений.</p> <p>Обновить соединение с репозиторием: выберите этот параметр, чтобы изменить схему, хранящуюся в репозитории, и решить, следует ли распространять изменения на все задания после завершения.</p> <p>Если вы просто хотите распространить изменения на текущее задание, вы можете выбрать «Нет» после завершения и снова выбрать метаданные этой схемы в окне «Содержимое репозитория».</p> <p>Обратите внимание, что если входное значение любого необнуляемого примитивного поля равно null, строка данных, включающая это поле, будет отклонена.</p> <p>Этот компонент предлагает преимущество функции динамической схемы. Это позволяет извлекать неизвестные столбцы из исходных файлов или копировать группы столбцов из источника без сопоставления каждого столбца по отдельности. Для получения дополнительной информации о динамических схемах см. Руководство пользователя Talend Studio.</p> <p>Эта функция динамической схемы разработана для извлечения неизвестных столбцов таблицы, и ее рекомендуется использовать только для этой цели; не рекомендуется использовать для создания таблиц.</p> <p>При использовании функции динамической схемы динамический столбец не содержит фактических имен столбцов входного файла. Если вы хотите, чтобы ваш выходной поток включал фактические имена столбцов,</p> |

| | |
|------------------------|--|
| | убедитесь, что ваш входной файл имеет строку заголовка и значение заголовка установлено правильно. |
| Skip empty rows | Установите этот флажок, чтобы пропустить пустые строки. |
| Uncompress as zip file | Установите этот флажок, чтобы распаковать входной файл. |
| Die on error | Установите флажок, чтобы остановить выполнение задания при возникновении ошибки. Снимите флажок, чтобы пропустить все строки с ошибкой и завершить процесс для безошибочных строк. Когда ошибки пропущены, вы можете собрать строки при ошибке, используя ссылку Строка > Отклонить . Чтобы поймать FileNotFoundException, вам также нужно установить этот флажок. |

№2



Преобразует и направляет данные из одного или нескольких источников в одно или несколько мест назначения.

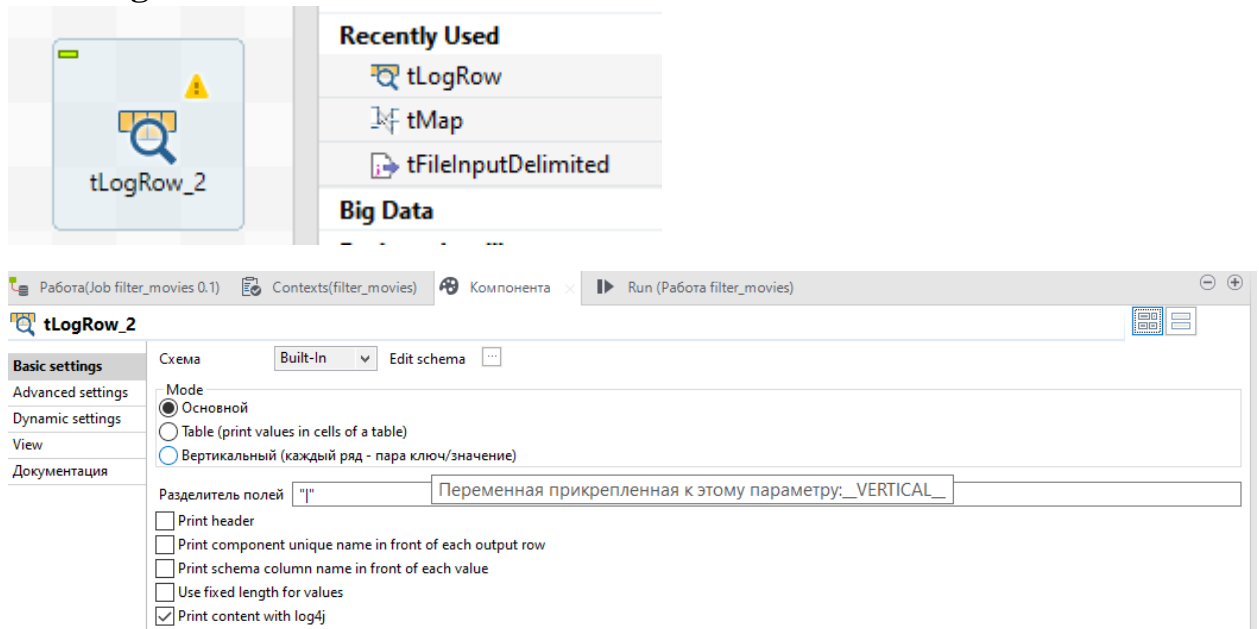
Подробное описание вкладок

| Вариант | Определение |
|------------|---|
| Map editor | Это позволяет вам определить свойства маршрутизации и преобразования tMap . При необходимости нажмите кнопку в верхней части области ввода, чтобы открыть диалоговое окно «Параметры свойств» , в котором доступны следующие параметры: Умирать при ошибке : установите этот флажок, если вы хотите завершить задание в случае ошибки. Этот флажок установлен по умолчанию. |

| | |
|--------------------------|--|
| | <p>Параллельный поиск : выберите этот флажок, чтобы максимизировать производительность преобразования данных в задании, которое обрабатывает несколько входных потоков поиска с большими объемами данных.</p> <p>Включить автоматическое преобразование типов : если ваши входные и выходные столбцы в сопоставлении имеют разные типы данных, установите этот флажок, чтобы включить автоматическое преобразование типов во время выполнения, чтобы избежать ошибок компиляции.</p> <p>Этот параметр включен по умолчанию, если установлен флажок «Включить автопреобразование типов» в представлении «Настройки проекта» при добавлении этого компонента. Вы также можете переопределить поведение преобразования по умолчанию для этого компонента, установив правила преобразования в представлении «Настройки проекта» . Для получения дополнительной информации см. Руководство пользователя Talend Studio .</p> <p>Обратите внимание, что автоматическое преобразование между Date и BigDecimal не поддерживается.</p> <p>Сохранить на диске : параметры, представленные в этой области, идентичны соответствующим параметрам, представленным на вкладках «Основные параметры» и «Дополнительные параметры» соответственно. Настройки, сделанные в диалоговом окне «Настройки свойств» , отражаются в соответствующих видах вкладок и наоборот.</p> <p>Этот компонент предлагает преимущество функции динамической схемы. Это позволяет извлекать неизвестные столбцы из исходных файлов или копировать группы столбцов из источника без сопоставления каждого столбца по отдельности. Для получения дополнительной информации о динамических схемах см. Руководство пользователя Talend Studio .</p> <p>Эта функция динамической схемы разработана для извлечения неизвестных столбцов таблицы, и ее рекомендуется использовать только для этой цели; не рекомендуется использовать для создания таблиц.</p> |
| Mapping links display as | <p>Авто : настройка по умолчанию кривые ссылки</p> <p>Кривые : отображение отображения в виде кривых</p> <p>Линии : сопоставление отображается в виде прямых линий. Эта последняя опция позволяет немного повысить производительность.</p> |
| Temp data directory path | <p>Введите путь, по которому вы хотите хранить временные данные, сгенерированные для загрузки поиска.</p> |

| | |
|---------|---|
| Preview | Предварительный просмотр — это мгновенный снимок данных Mapper. Он становится доступным, когда свойства Mapper заполнены данными. Синхронизация предварительного просмотра вступает в силу только после сохранения изменений. |
|---------|---|

№3 tLogRow



Отображает данные или результаты в консоли «Выполнение» для отслеживания обработанных данных.

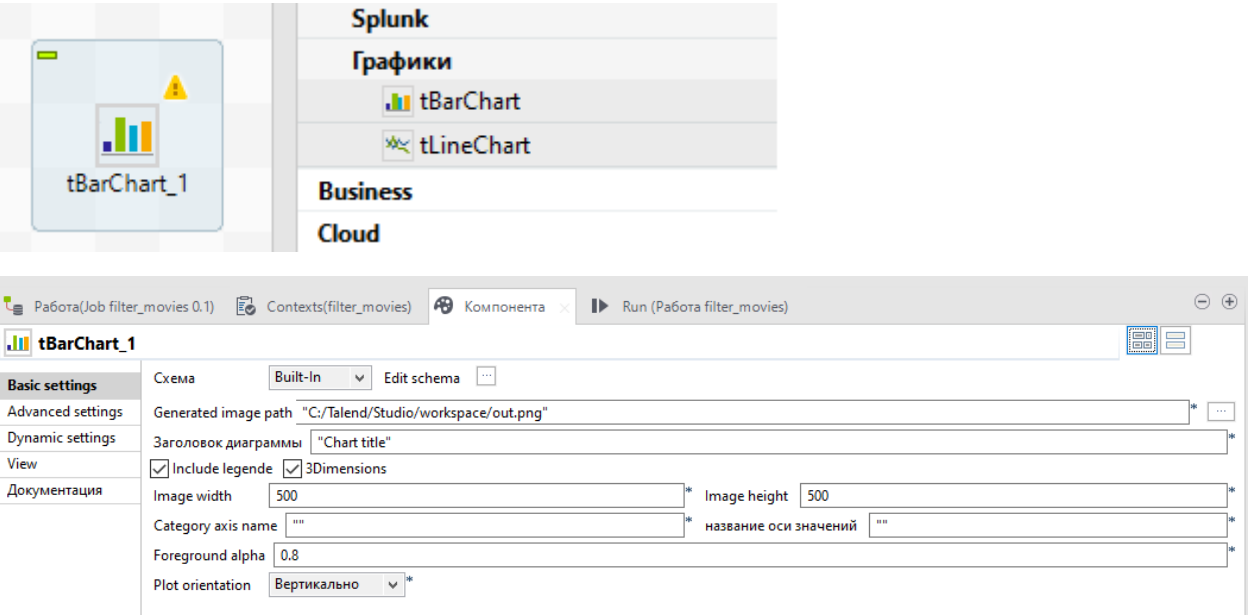
Подробное описание вкладок

| Вариант | Определение |
|------------------------|---|
| Schema and Edit schema | <p>Схема представляет собой описание строки. Он определяет количество полей (столбцов), которые необходимо обработать и передать следующему компоненту. При создании задания Spark избегайте зарезервированного слова <code>line</code> при именовании полей. Щелкните Изменить схему , чтобы внести изменения в схему. Если текущая схема имеет тип Репозиторий , доступны три варианта:</p> <p>Просмотреть схему : выберите этот параметр, чтобы просмотреть только схему.</p> <p>Изменить на встроенное свойство : выберите этот параметр, чтобы изменить схему на встроенную для локальных изменений.</p> <p>Обновить соединение с репозиторием : выберите этот параметр, чтобы изменить схему, хранящуюся в репозитории, и решить, следует ли распространять изменения на все задания после завершения.</p> <p>Если вы просто хотите распространить изменения на текущее задание, вы можете выбрать «Нет» после завершения и снова</p> |

| | |
|---|---|
| | <p>выбрать метаданные этой схемы в окне «Содержимое репозитория».</p> <p>Этот компонент предлагает преимущество функции динамической схемы. Это позволяет извлекать неизвестные столбцы из исходных файлов или копировать группы столбцов из источника без сопоставления каждого столбца по отдельности. Для получения дополнительной информации о динамических схемах см. Руководство пользователя Talend Studio .</p> <p>Эта функция динамической схемы разработана для извлечения неизвестных столбцов таблицы, и ее рекомендуется использовать только для этой цели; не рекомендуется использовать для создания таблиц.</p> |
| Sync columns | Нажмите, чтобы синхронизировать схему выходного файла со схемой входного файла. Функция синхронизации доступна только тогда, когда компонент связан с предыдущим компонентом с помощью соединения Row . |
| Basic | Отображает выходной поток в базовом режиме. |
| Table | Отображает поток вывода в ячейках таблицы. |
| Vertical | <p>Отображает каждую строку выходного потока в виде списка "ключ-значение".</p> <p>Выбрав этот режим, вы можете выбрать отображение либо уникального имени, либо метки компонента, либо того и другого для каждой выходной строки.</p> |
| Separator | Введите разделитель, который будет разделять данные на дисплее журнала. |
| Print header | Установите этот флажок, чтобы включить заголовок входящего потока в выходной дисплей. |
| Print component unique name in front of each output row | Установите этот флажок, чтобы показать уникальное имя компонента перед каждой строкой вывода, чтобы различать выходные данные в случае использования нескольких компонентов tLogRow . |
| Print schema column name in front of each value | Установите этот флажок, чтобы получить метки столбцов из выходной схемы. |
| Use fixed length for values | Установите этот флажок, чтобы установить фиксированную ширину для отображения значений. |

| | |
|--------------------------|--|
| Print content with log4j | Установите этот флажок для вывода содержимого потока данных, если функция log4j активирована и уровень вывода log4j установлен на INFO или ниже, чем INFO для вашего задания. Чтобы узнать, как активировать функцию log4j и настроить уровень вывода log4j, обратитесь к Руководству пользователя Talend Studio . |
|--------------------------|--|

№4 tBarChart



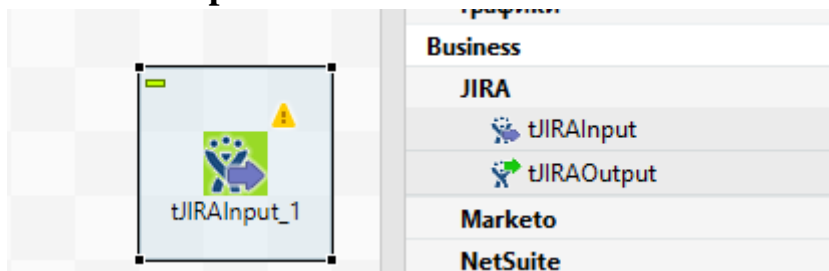
Создает гистограмму из входных данных для облегчения технического анализа.

Подробное описание вкладок

| Вариант | Определение |
|------------------------|--|
| Schema and Edit schema | <p>Схема представляет собой описание строки. Он определяет количество полей (столбцов), которые необходимо обработать и передать следующему компоненту. При создании задания Spark избегайте зарезервированного слова line при именовании полей.</p> <p>Щелкните Изменить схему , чтобы внести изменения в схему. Если текущая схема имеет тип Репозиторий , доступны три варианта:</p> <p>Просмотреть схему : выберите этот параметр, чтобы просмотреть только схему.</p> <p>Изменить на встроенное свойство : выберите этот параметр, чтобы изменить схему на встроенную для локальных изменений.</p> <p>Обновить соединение с репозиторием : выберите этот параметр, чтобы изменить схему, хранящуюся в</p> |

| | |
|--|---|
| | <p>репозитории, и решить, следует ли распространять изменения на все задания после завершения.</p> <p>Если вы просто хотите распространить изменения на текущее задание, вы можете выбрать «Нет» после завершения и снова выбрать метаданные этой схемы в окне «Содержимое репозитория».</p> |
| Sync columns | Нажмите, чтобы синхронизировать схему выходного файла со схемой входного файла. Функция синхронизации отображается только после того, как соединение Row связано с выходным компонентом. |
| Generated image path | Имя и путь к выходному файлу изображения. |
| Chart title | Введите название гистограммы, которую необходимо сгенерировать. |
| Include legend | Установите этот флажок, если вы хотите, чтобы линейчатая диаграмма включала легенду, обозначающую все ряды разными цветами. |
| 3Dimensions | Установите этот флажок, чтобы создать изображение с 3D-эффектом. По умолчанию этот флажок установлен, и столбцы, представляющие серию каждой категории, будут располагаться друг над другом. Если этот флажок снят, будет создано 2D-изображение с полосами, расположенными одна за другой вдоль оси категорий. |
| Image width and Image height | Введите ширину и высоту файла изображения в пикселях. |
| Category axis name and Value axis name | Введите имя оси категорий и имя оси значений. |
| Foreground alpha | Введите целое число в диапазоне от 0 до 100, чтобы задать прозрачность изображения. Чем меньше введенное число, тем более прозрачным будет изображение. |
| Plot orientation | Выберите ориентацию гистограммы |

№5 tJIRAInput

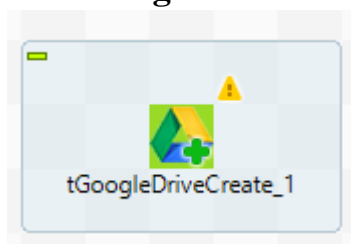


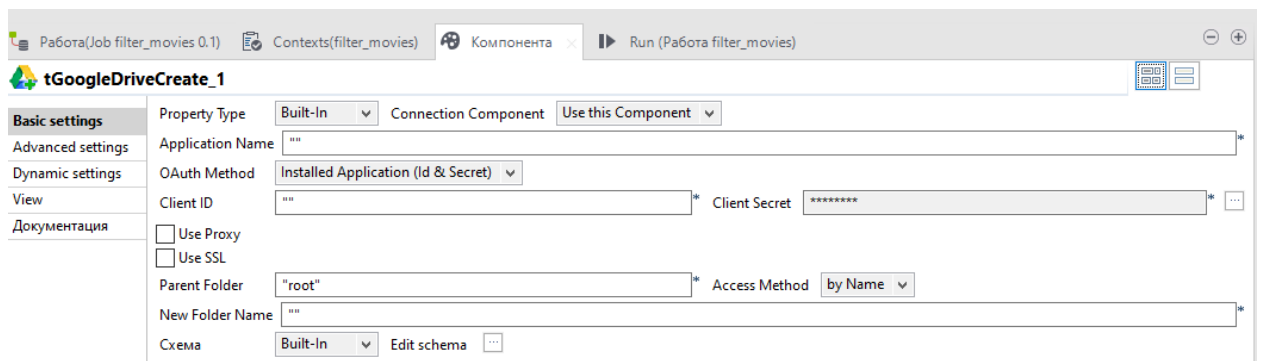
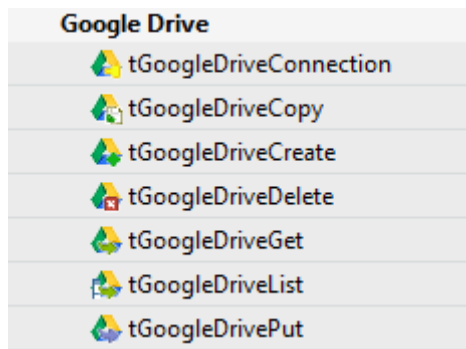
Извлекает информацию о проблеме на основе запроса JQL или извлекает информацию о проекте на основе указанного идентификатора проекта из JIRA.

Подробное описание вкладок

| Вариант | Определение |
|------------------------|---|
| User Id and Password | Укажите данные аутентификации пользователя, используемые для доступа к JIRA. Чтобы ввести пароль, нажмите кнопку [...] рядом с полем пароля, а затем во всплывающем диалоговом окне введите пароль в двойных кавычках и нажмите кнопку ОК , чтобы сохранить настройки. |
| JIRA resource | В раскрывающемся списке выберите ресурс JIRA: ISSUE или PROJECT . |
| Schema and Edit schema | Схема представляет собой описание строки. Он определяет количество полей (столбцов), которые необходимо обработать и передать следующему компоненту. Обратите внимание, что схема этого компонента доступна только для чтения с одним столбцом json типа String . Он хранит информацию о проблеме или проекте в формате JSON. Вы можете нажать Изменить схему , чтобы просмотреть схему. |
| JQL | Укажите запрос JQL (JIRA Query Language), на основе которого вы будете искать задачи в JIRA. Это поле доступно только в том случае, если в раскрывающемся списке ресурсов Jira выбран пункт ПРОБЛЕМА . |
| Project ID | Введите идентификатор проекта, информацию о котором вы хотите получить из JIRA. Это поле доступно, только если в раскрывающемся списке ресурсов Jira выбран ПРОЕКТ . |

№6 tGoogleDriveCreate





Создает новую папку на Google Диске.

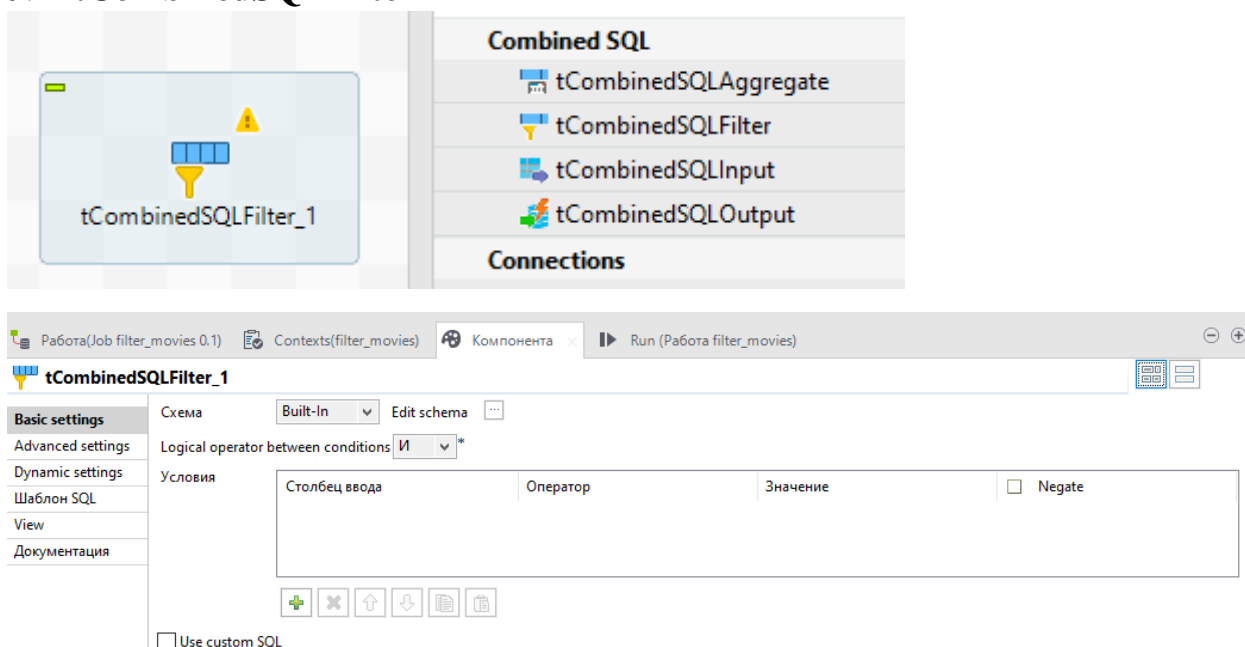
Подробное описание вкладок

| Вариант | Определение |
|----------------------|---|
| Property Type | <p>Выберите способ установки сведений о соединении.</p> <p>Встроенный : Детали подключения будут установлены локально для этого компонента. Вам необходимо указать значения для всех связанных свойств соединения вручную.</p> <p>Репозиторий : сведения о соединении, хранящиеся централизованно в Репозиторий > Метаданные , будут повторно использоваться этим компонентом. Вам нужно нажать кнопку [...] рядом с ним и во всплывающем диалоговом окне «Содержимое репозитория» выбрать детали подключения для повторного использования, и все связанные свойства подключения будут автоматически заполнены.</p> <p>Это свойство недоступно, если в раскрывающемся списке Компонент соединения выбран другой компонент соединения .</p> |
| Connection Component | <p>Выберите компонент, который открывает соединение с базой данных для повторного использования этим компонентом.</p> |
| Application Name | <p>Имя приложения, необходимое Google Диску для доступа к его API.</p> |
| OAuth Method | <p>Выберите метод OAuth, используемый для доступа к Google Диску, из раскрывающегося списка.</p> <p>Токен доступа (устаревший) : использует токен доступа для доступа к Google Диску.</p> <p>Установленное приложение (идентификатор и секрет) : использует идентификатор клиента и секрет клиента,</p> |

| | |
|-----------------------------|--|
| | <p>созданные с помощью Google API Console, для доступа к Google Диску. Дополнительные сведения об этом методе см. в разделе Google Identity Platform > Установленные приложения . Установленное приложение (JSON) : использует файл JSON с секретом клиента, созданный с помощью Google API Console и содержащий идентификатор клиента, секрет клиента и другие параметры OAuth 2.0 для доступа к Google Диску.</p> <p>Учетная запись службы : использует файл JSON учетной записи службы, созданный с помощью Google API Console, для доступа к Google Диску. Дополнительные сведения об этом методе см. в разделе Google Identity Platform > Сервисные аккаунты .</p> <p>Для получения более подробной информации о том, как получить доступ к Google Диску с помощью каждого метода, см. раздел Методы OAuth для доступа к Google Диску .</p> |
| Access Token | <p>Токен доступа, созданный с помощью Google Developers OAuth 2.0 Playground.</p> <p>Это свойство доступно, только если Access Token оно выбрано в раскрывающемся списке Метод OAuth .</p> |
| Client ID and Client Secret | <p>Идентификатор клиента и секрет клиента.</p> <p>Эти два свойства доступны, только если Installed Application (Id & Secret) выбрано в раскрывающемся списке Метод OAuth .</p> |
| Client Secret JSON | <p>Путь к секретному JSON-файлу клиента.</p> <p>Это свойство доступно, только если Installed Application (JSON) оно выбрано в раскрывающемся списке Метод OAuth .</p> |
| Service Account JSON | <p>Путь к JSON-файлу сервисного аккаунта.</p> <p>Это свойство доступно, только если Service Account оно выбрано в раскрывающемся списке Метод OAuth .</p> |
| Use Proxy | <p>Установите этот флажок, если вы работаете за прокси-сервером. При установленном флажке необходимо указать значение для следующих параметров:</p> <p>Хост : IP-адрес прокси-сервера HTTP.</p> <p>Порт : номер порта прокси-сервера HTTP.</p> |
| Use SSL | <p>Установите этот флажок, если для доступа к Google Диску используется соединение SSL. При установленном флажке необходимо указать значение для следующих параметров:</p> <p>Алгоритм : название алгоритма шифрования SSL.</p> <p>Файл хранилища ключей : путь к файлу хранилища сертификатов, который содержит список сертификатов, которым доверяет клиент.</p> <p>Пароль : пароль, используемый для проверки целостности данных TrustStore.</p> |
| Parent Folder | <p>Имя или идентификатор родительской папки, в которой будет создана новая папка.</p> |

| | |
|------------------------|---|
| Access Method | Выберите способ указания родительской папки: либо , by Nameлибо by Id. |
| New Folder Name | Имя новой создаваемой папки. |
| Schema and Edit schema | <p>Схема представляет собой описание строки и определяет поля, которые необходимо обработать и передать следующему компоненту.</p> <p>Схема этого компонента доступна только для чтения. Вы можете нажать кнопку [...] рядом с Изменить схему , чтобы просмотреть предопределенную схему, которая содержит следующие поля:</p> <p>parentFolderId : идентификатор родительской папки. newFolderId : идентификатор новой папки.</p> |

№7 tCombinedSQLFilter



The image displays the tCombinedSQLFilter component in a software interface. The top section shows a visual representation of the component on a canvas and a list of components: tCombinedSQLAggregate, tCombinedSQLFilter, tCombinedSQLInput, and tCombinedSQLOutput. The bottom section shows the configuration window for tCombinedSQLFilter_1, including settings for schema, logical operator, and conditions.

Фильтрует данные путем реорганизации, удаления или добавления столбцов на основе исходной таблицы и для фильтрации данного источника данных с использованием условий фильтрации.

tCombinedSQLFilter позволяет изменять схему исходной таблицы посредством сопоставления имен столбцов и определять фильтр строк в этой таблице. Поэтому его можно использовать для одновременной фильтрации столбцов и строк. Этот компонент имеет возможности работы в реальном времени, так как он запускает фильтрацию данных на самой СУБД.






Подробное описание вкладок

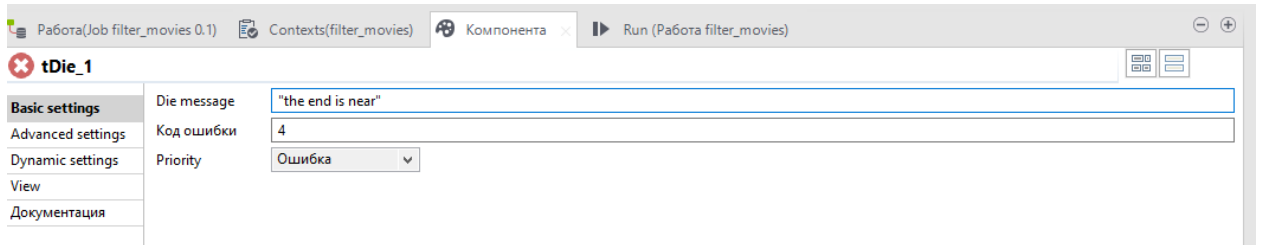
| Вариант | Определение |
|---------|-------------|
|---------|-------------|

| | |
|-------------------------------------|---|
| Schema and Edit schema | <p>Схема представляет собой описание строки. Он определяет количество полей (столбцов), которые необходимо обработать и передать следующему компоненту. При создании задания Spark избегайте зарезервированного слова line при именовании полей.</p> <p>Щелкните Изменить схему , чтобы внести изменения в схему. Если текущая схема имеет тип Репозиторий , доступны три варианта:</p> <p>Просмотреть схему : выберите этот параметр, чтобы просмотреть только схему.</p> <p>Изменить на встроенное свойство : выберите этот параметр, чтобы изменить схему на встроенную для локальных изменений.</p> <p>Обновить соединение с репозиторием : выберите этот параметр, чтобы изменить схему, хранящуюся в репозитории, и решить, следует ли распространять изменения на все задания после завершения. Если вы просто хотите распространить изменения на текущее задание, вы можете выбрать « Нет » после завершения и снова выбрать метаданные этой схемы в окне «Содержимое репозитория ».</p> <p>Нажмите «Синхронизировать столбцы» , чтобы получить схему из предыдущего компонента, подключенного к заданию.</p> |
| Logical operator between conditions | Выберите логический оператор между условиями фильтрации, определенными на панели «Условия» . |
| Conditions | Выберите тип предложения WHERE вместе со значениями и столбцами, которые будут использоваться для фильтрации строк. |
| Use custom SQL | Настройте предложение WHERE, установив этот флажок и отредактировав поле SQL Condition . |

№8 tDie

Журналы и ошибки

| |
|---|
|  tAssert |
|  tAssertCatcher |
|  tChronometerStart |
|  tChronometerStop |
|  tDie |



Запускает компонент tLogCatcher для полного журнала перед завершением задания.

Компоненты tDie и tWarn тесно связаны с компонентом tLogCatcher. Как правило, их целесообразно использовать вместе с tLogCatcher, чтобы инкапсулировать собранные данные журнала и передать их на определенный выход.

Этот компонент выдает ошибку и убивает задание. Если вы просто хотите выдать предупреждение, см. документацию по tWarn.

Подробное описание вкладок

| Вариант | Определение |
|------------|---|
| Error code | Введите сообщение, которое будет отображаться перед уничтожением задания. |
| Error code | При необходимости введите код ошибки в виде целого числа. |
| Priority | Установите уровень приоритета в виде целого числа |