

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

Башкатова Анна Денисовна

ОТЧЕТ

по дисциплине «Инструменты для хранения и обработки больших данных»

Тема: «03 Архитектура хранилищ данных: традиционная и облачная»

Направление подготовки (специальность) 38.03.05 – бизнес-информатика

Направленность (профиль) образовательной программы «Аналитика данных
и эффективное управление»

Курс обучения: 3

Форма обучения: очная

Руководитель: Босенко Т. М.

Москва
2023

В данной практической работе необходимо описать 8-10 компонентов в ETL Pentaho DI и Talend.

Практика 3

В качестве практики вам необходимо выявить 8-10 подсистем в

ETL Pentaho DI и Talend,

написать отчет, в котором Вы приложите **print screen** компонента (ETL подсистемы) и напишите про его свойства. Результат сохраните в вашем **Git**.

Перечень описываемых компонентов:

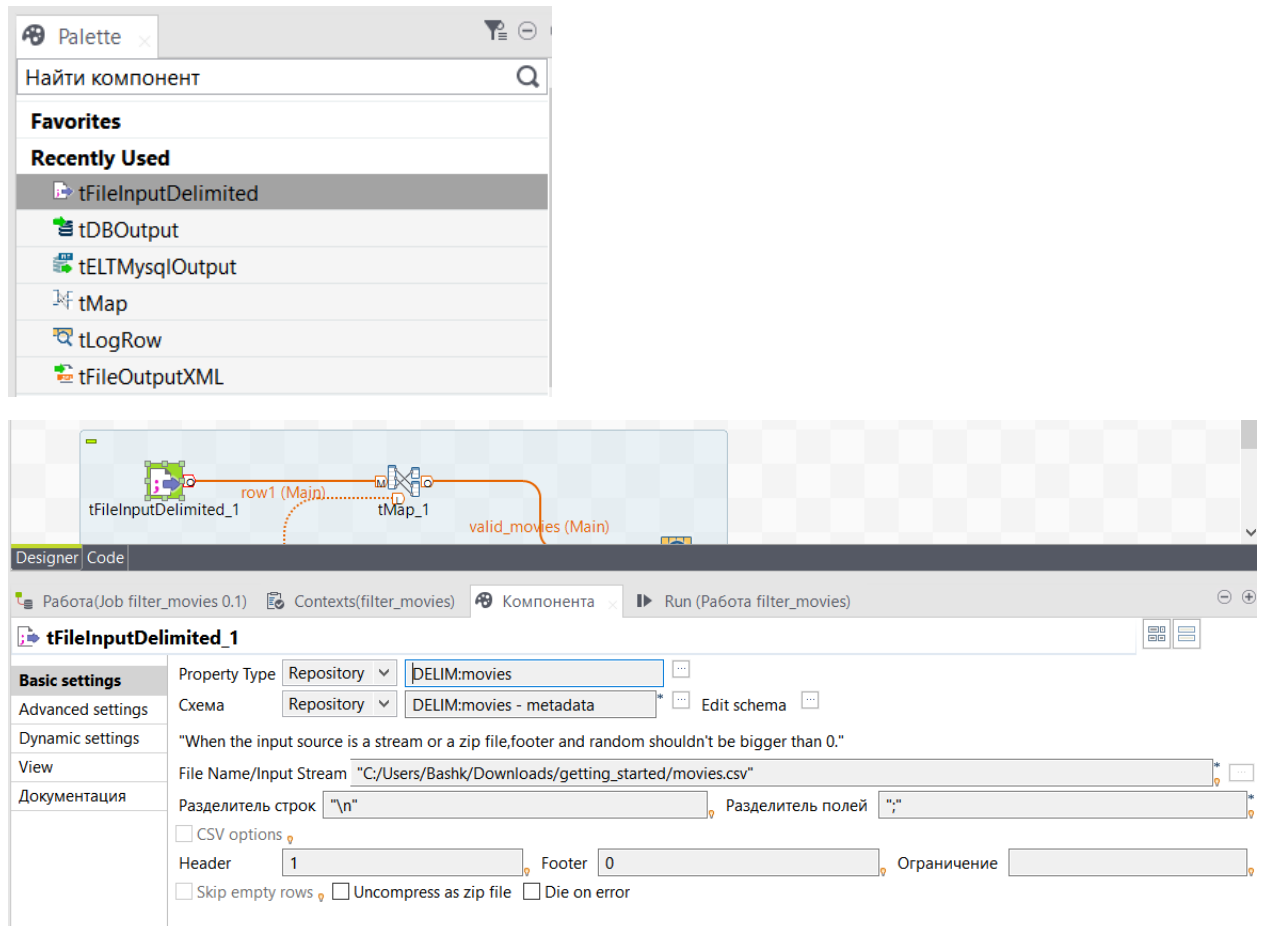
Talend	3
1. tFileinputdelimited	3
2. tmap	4
3. tLogRow	5
4. tDBOutput	7
5. tdie	8
6. tSendMail	8
7. tJava	10
8. tAggregateRow	10
Pentaho DI	14
1. Microsoft Excel input	14
2. Sort rows	14
3. Merge join	15
4. Select values	16
5. Text file output	16
6. Table output	17
7. Write to log	18
8. Number range	19

Talend

Talend - это инструмент ETL, который используется для интеграции данных. Talend предоставляет решение для подготовки данных, качества данных, интеграции данных и больших данных.

1. tFileinputdelimited

Этот компонент используется для чтения любых файлов с разделителями. Считывает заданный файл строка за строкой с простыми разделенными полями.



Эти свойства используются для настройки tFileInputDelimited, работающего в платформе стандартных заданий. Стандартный компонент tFileInputDelimited относится к семейству File. Компонент в этом фреймворке доступен во всех Talend продукция.

Property Type: Built-in – встроенный – данные о свойствах не хранятся централизованно; *repository* – репозиторий – выберите файл репозитория, в котором Свойства сохраняются.

Имя файла/поток: Имя файла – имя и путь к обрабатываемому файлу. Поток – Поток данных, подлежащих обработке. Данные должны быть добавлены в поток, чтобы tFileInputDelimited мог получить эти данные с помощью соответствующей репрезентативной переменной.

Разделитель строк – разделитель, используемый для идентификации конца строки.

Разделитель полей – введется символ, строка или регулярное выражение для разделения поля для передаваемых данных.

Параметры CSV: 1) Escape-символ: введите escape-символ между double кавычки. 2) Текстовое вложение: введите символ вложения (только один символ) между двойными кавычками. Например, необходимо вводить, когда в качестве вложения используются двойные кавычки (") характер.

Заголовок вводится количество строк, которые нужно пропустить в начале файла.

Нижний колонтитул – количество строк, которые должны быть пропущены в конце файла.

Предел – максимальное количество обрабатываемых строк. Если Limit = 0, ни одна строка не считывается и не обрабатывается.

Пропуск пустых строк – устанавливается этот флажок, чтобы пропустить пустые строки.

2. tmap

Компонент **tmap** – этот компонент помогает преобразовывать данные с помощью некоторых встроенных функций, таких как поиск, объединения и т. д.

The screenshot displays the Talend Open Studio interface for configuring the **tMap** component. The top section shows the 'Basic settings' tab with 'Mapping links display as: Auto'. The middle section shows the 'Map Editor' with two input rows (row1 and row2) and a 'valid_movies' output row. The bottom section shows the 'Schema editor' with two tables: 'row1' and 'valid_movies'.

Колонка	К...	Тип	✓ N.	Date Pattern (Ctrl+...)	Длина	Precision	Default	Коммента...
movieID		Integer	✓		4	0		
title		String	✓		72	0		
releaseYear		Integer	✓		4	0		
url		String	✓		120	0		
directorID		String	✓		3	0		

Колонка	К...	Тип	✓ N.	Date Pattern (Ctrl+...)	Длина	Precision	Default	Коммента...
movieID		Integer	✓		4	0		
title		String	✓		72	0		
directorBy		String	✓		20	0		
releaseYear		Integer	✓		4	0		
url		String	✓		120	0		

tMap - это все в одном компоненте. Все процессы происходят в Редакторе карт в компоненте tMap. Чтобы открыть Редактор карт, дважды щелкните компонент tMap или В разделе В основных настройках откройте окно Редактор карт. Несколько панелей находятся в редакторе карт. Компонент tMap не является началом и концом задания. tMap должен иметь как входной, так и выходной компонент.

Будет только одно подключение к главной строке, а остальные все входные таблицы будут рассматриваться как соединение поиска. Строки подстановки взяты из вторичных потоков данных. Эти справочные данные могут прямо или косвенно зависеть от первичного потока.

Имена таблиц ввода/вывода в Редакторе карт совпадают с именами соединений строк входящих и исходящих потоков.

Мы можем предварительно просмотреть редактор карт в основных настройках.

Редактор карт имеет следующие панели:

- Панель ввода
- Вариативная панель
- Панель поиска
- Панель «Вывод»
- Редактор схем
- Редактор выражений

Сопоставление ссылок отображается как 1) Авто: настройка по умолчанию — кривые дюны 2) Кривые: отображение отображения в виде кривых 3) Линии: отображение отображается как прямое обводки. Этот последний вариант позволяет немного повысить производительность.

Путь к каталогу временных данных вводится путь, по которому вы хотите сохранить временные данные генерируется для загрузки подстановки. Дополнительные сведения об этой папке см. в разделе Talend Studio User Руководство.

Предварительный просмотр представляет собой мгновенный снимок данных Mapper. Он становится доступно, когда свойства Mapper были заполнены данными. Предварительная синхронизация вступает в силу только после сохранения изменений.

3. tLogRow

The screenshot displays the Talend Studio interface with a workflow diagram and the configuration for the tLogRow component.

Workflow Diagram: The diagram shows a sequence of components: tFileInputDelimited_1, tMap_1, valid_moves (Main), and tLogRow_1. Data flows from tFileInputDelimited_1 to tMap_1, then to valid_moves (Main), and finally to tLogRow_1. There is also a 'directors' component at the bottom left.

tLogRow_1 Configuration: The configuration panel for tLogRow_1 is visible, showing the 'Basic settings' tab. The 'Mode' is set to 'Основной' (Basic). The 'Title printing mode' is set to 'Print unique name and label'. The 'Print content with log4j' checkbox is checked.

Component Palette: The right-hand side of the interface shows the 'Component Palette' with various components categorized under 'Big Data', 'Business Intelligence', 'DB', 'Cloud', and 'Other'. The 'tLogRow' component is highlighted in the 'Big Data' category.

Отображение данных или результатов в консоли «Выполнить» для мониторинга обрабатываемых данных. Стандартный компонент tLogRow относится к семейству Logs & Errors.

Схема — это описание строки. Он определяет количество полей (столбцы), которые должны быть обработаны и переданы следующему компоненту. При создании Spark Job, избегайте зарезервированного слова при именовании Поля.line Нажатие «Изменить» схему, чтобы внести изменения в схему. Если текущая схема относится к типу Repository, доступны три варианта:

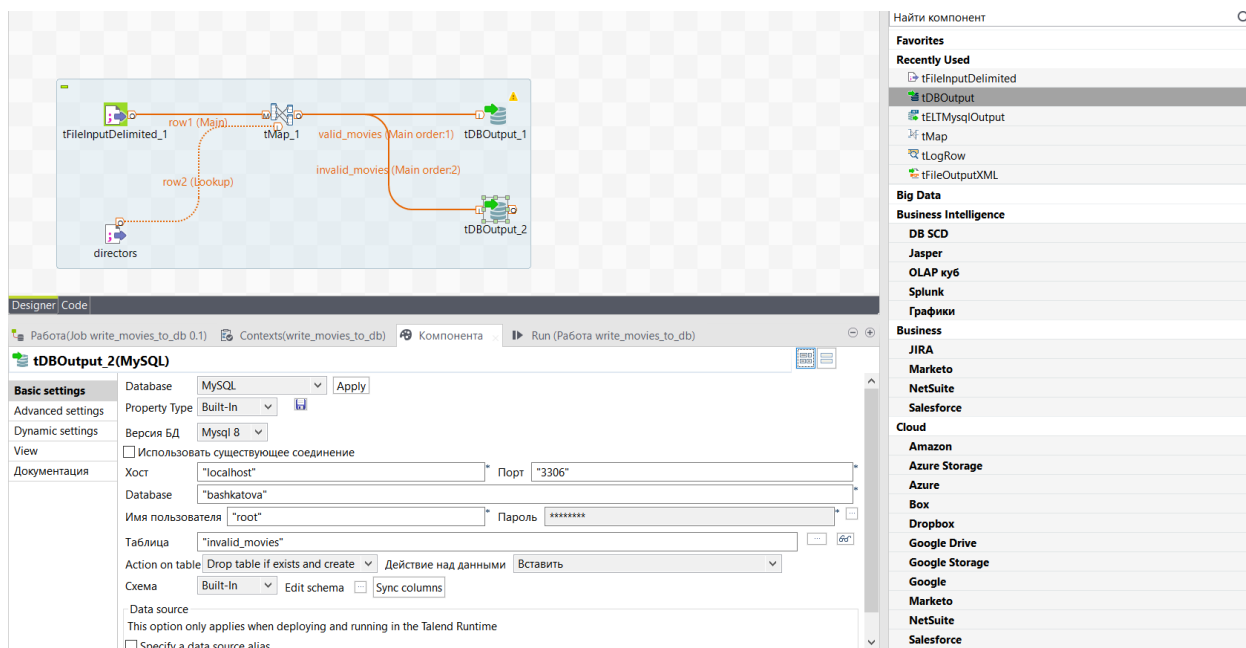
- Схема просмотра: выберите эту схему для просмотра только схемы.
- Изменение встроенного свойства: выберите этот параметр, чтобы изменить схему на Встроенная для локальных изменений.
- Подключение к репозиторию обновлений: Выберите этот параметр, чтобы изменить схему, хранящуюся в репозитории, и решить, следует ли распространять изменения на все задания после их завершения.

Если нужно распространить изменения на текущее задание, можно выбрать «Нет» по завершении и выбрать метаданные этой схемы снова в окне "Содержимое репозитория".

Этот Компонент предлагает преимущество функции динамической схемы. Это позволяет: Извлечение неизвестных столбцов из исходных файлов или копирование пакетов столбцов из источника без сопоставления каждого столбца по отдельности.

Синхронизация столбцов	Нажмите, чтобы синхронизировать схему выходного файла с входными данными Схема файла. Функция синхронизации доступна только в том случае, если компонент связан с предыдущим компонентом с помощью подключения Row .
Основной	Отображает выходной поток в базовом режиме.
Стол	Отображает выходной поток в ячейках таблицы.
Вертикальный	Отображает каждую строку выходного потока в виде списка "ключ-значение". Если выбран этот режим, вы можете выбрать отображение одного из уникальных имен или метка компонента, или и то, и другое, для каждой выходной строки.
Разделитель	Введите разделитель, который будет разграничивать данные на дисплее журнала.
Печать заголовка	Установите этот флажок, чтобы включить заголовок входных данных поток на выходном дисплее.
Печать уникального имени компонента перед каждым Выходная строка	Установите этот флажок, чтобы отобразить уникальное имя компонента перед ним каждой выходной строки для дифференциации выходных данных в случае использования нескольких компонентов tLogRow .
Печать имени столбца схемы перед каждым из них ценность	Установите этот флажок, чтобы получить метки столбцов из выходных данных схема.
Использование фиксированной длины для значения	Установите этот флажок, чтобы задать фиксированную ширину для значения дисплей.

4. tDBOutput



Записывает, обновляет, вносит изменения или подавляет записи в базе данных. Этот компонент работает с различными базами данных в зависимости от выбора.

Этот компонент служит точкой входа для следующих баз данных. Чтобы настроить этот компонент, выберите тип базы данных из списка База данных и нажмите кнопку Применить к ее Базовой. Вид настроек. Дополнительные сведения о конкретных свойствах базы данных см. соответствующую документацию. Например, для получения сведений о том, как настроить tDBCDC для работы с AS400 см. документацию tAS400CDC.

Доступ; Амазонка Аврора; Amazon Mysql; Амазонка Оракул; Красное смещение Амазонки; AC400; Озеро Дельта; Экзасол; Жар-птица; Зеленая слива; IBM DB2; Информикс; Энгр; Интербаза; JDBC; MemSQL; Microsoft SQL Server; MySQL; Netezza; Оракул; ПарАксель; PostgreSQL; ПостгресПлюс; САФана; Единый магазин; Снежинка; SQLite; Sybase (ASE и IQ); Терадата; Векторный мудрый; Вертика.

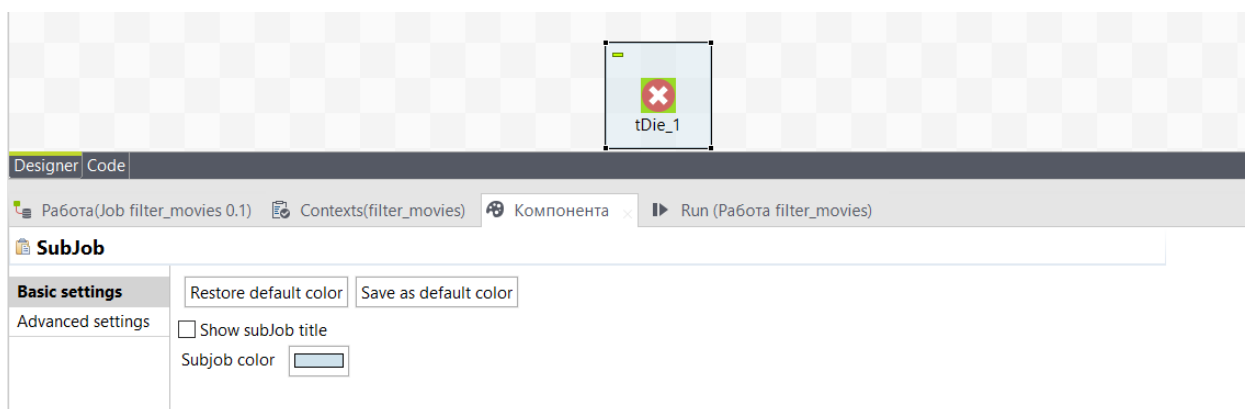
Property Type. Тип свойства позволяет вам указать, должны ли параметры подключения к вашей базе данных быть взяты из репозитория (как в нашем примере) или встроены в сам компонент. Как обсуждалось в разделе Работа с базами данных, я бы рекомендовал вам не использовать этот параметр. Добавьте компонент подключения t[DB] и используйте его для поддержания подключений к базе данных.

Используйте существующее соединение Этот параметр позволяет вам использовать существующее подключение к базе данных. То есть соединение, которое уже было установлено с помощью компонента подключения t[DB]. При выборе этого параметра отобразится новый выпадающий список. В этом списке будут показаны все доступные подключения. Это способ подключения, который я бы всегда рекомендовал.

Стол Здесь вы можете указать имя таблицы, с которой будут выполняться операции как с языком определения данных (DDL), так и с языком манипулирования данными (DML).

Действие на столе Этот параметр позволяет вам указать любую операцию языка определения данных (DDL), которая должна быть выполнена с таблицей. Доступны следующие опции.

5. tdie



Эти свойства используются для настройки tDie, работающего в среде Standard Job. Стандартный компонент tDie относится к семейству Logs & Errors. Компонент в этом фреймворке доступен во всех Talend продукция.

Основные настройки

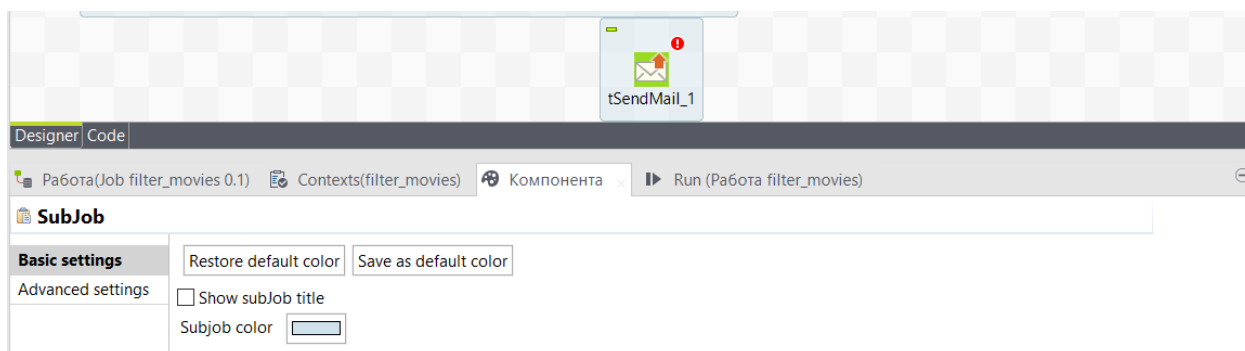
Сообщение о кубке - Введите сообщение, которое будет отображаться перед завершением задания.

Код ошибки - При необходимости введите код ошибки в виде целого числа.

Заметка: обратите внимание, что любое значение, превышающее 255, не может быть использовано в качестве код ошибки в Linux.

Приоритет - Установите уровень приоритета в виде целого числа

6. tSendMail



Основные настройки

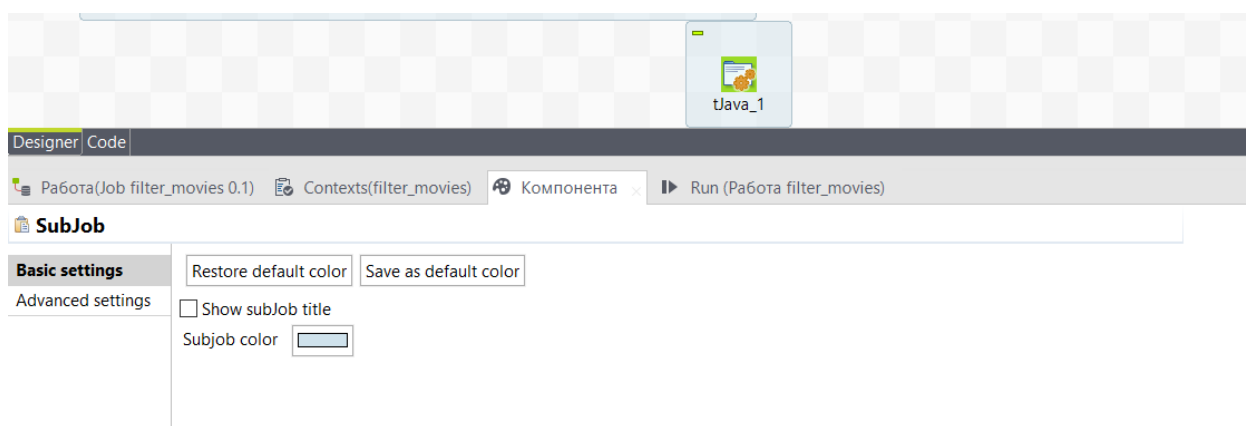
Кому	Основной адрес электронной почты получателя.
От	Адрес электронной почты сервера отправки.

Кому	Основной адрес электронной почты получателя.
Показать имя отправителя	Установите этот флажок, если вы хотите Имя отправителя, которое будет отображаться в сообщениях.
Кубовый	Адреса электронной почты вторичных получателей сообщения электронной почты, направленного на другое.
Скрытая копия	Адреса электронной почты вторичных получателей сообщения электронной почты. Получатели, указанные в поле « Скрытая копия », получают копию сообщения, но не отображаются ни на одной другой копии получателя.
Тема	Заголовок письма.
Сообщение	Основное сообщение электронного письма. Нажмите клавиши CTRL+ПРОБЕЛ , чтобы Отображение списка доступных переменных.
Умереть, если файл вложения не существовать	Этот флажок установлен по умолчанию. Снимите этот флажок, если вы хотите, чтобы сообщение было отправлено, даже если оно не являются вложениями.
Вложения	<p>Укажите прикрепленные и отправленные файлы вместе с сообщением электронной почты. Нажмите кнопку [+], чтобы добавить столько строк, сколько необходимо и задайте значения следующих столбцов для каждой строки:</p> <ul style="list-style-type: none">• Файл: введите путь к файлу Вложенный файл или путь к папке, если все файлы Непосредственно под ним будет крепиться.• Кодирование передачи контента: Щелкните ячейку и выберите значение кодировки передачи содержимого из списка.
Другие заголовки	Нажмите кнопку с плюсом, чтобы добавить как можно больше строк по мере необходимости, где вы можете ввести ключ и соответствующее значение любого заголовка информация, не относящаяся к стандартному заголовку.
Хост и порт SMTP	IP-адрес SMTP-сервера, используемого для отправки электронные письма.
Поддержка SSL	Установите этот флажок, чтобы проверить подлинность сервер на стороне клиента по протоколу SSL.
Поддержка STARTTLS	Установите этот флажок, чтобы проверить подлинность сервер на стороне клиента по протоколу STARTTLS.
Важность	Выберите в списке уровень приоритета Ваши сообщения.

Кому	Основной адрес электронной почты получателя.
Режим аутентификации	<p>Выберите один из трех режимов аутентификации из выпадающего списка.</p> <ul style="list-style-type: none"> • Нет аутентификации, которая не требует аутентификации. • Basic, который проверяет подлинность доступа с помощью имени пользователя и пароль. Вам необходимо указать имя пользователя и пароль для отправки почта. Чтобы ввести пароль, нажмите кнопку [...] рядом с полем Пароль введите пароль в поле двойные кавычки во всплывающем диалоговом окне и нажмите кнопку «ОК», чтобы сохранить настройки. • OAuth2, который проверяет подлинность доступа с помощью OAuth 2.0. Для отправки файла необходимо указать имя пользователя и токен доступа OAuth. почта. Чтобы ввести маркер доступа OAuth, нажмите кнопку [...] рядом с доступом OAuth Token, введите токен доступа OAuth в двойных кавычках в всплывающее диалоговое окно и нажмите кнопку «ОК», чтобы сохранить файл Параметры.
Умри на ошибке	Этот флажок установлен по умолчанию. Снимите флажок, чтобы пропустить строку об ошибке и завершить процесс для безошибочных строк.

7. tJava

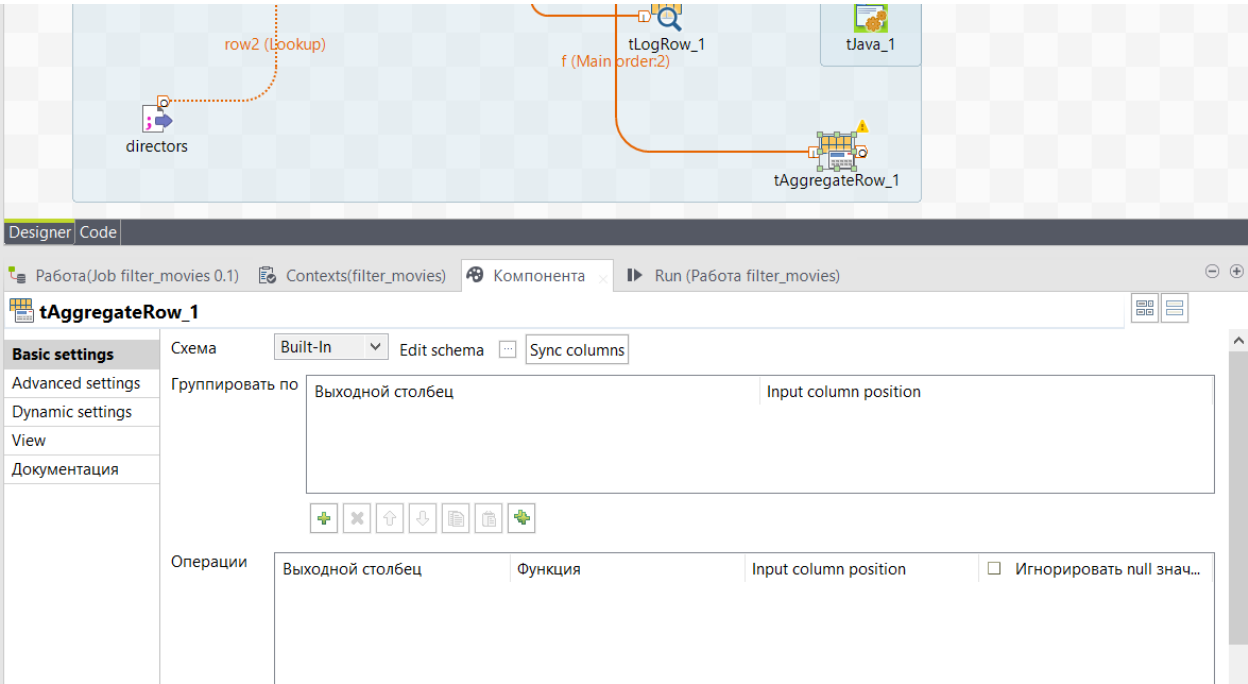
Расширяет функциональные возможности задания Talend с помощью пользовательской Java Команды. tJava позволяет вводить персонализированные для того, чтобы интегрировать его в программу Talend. Вы можете выполнить Этот код только один раз.



Основные настройки

Код	Введите код Java, который вы хотите выполнить в соответствии с задачей Нужно выполнять. Для получения дополнительной информации о синтаксисе функций Java Для Talend см. раздел Содержание справки Talend Studio (Справка > Руководство разработчика > Справочник по API).
------------	--

8. tAggregateRow



Получает поток и агрегирует его на основе одного или нескольких столбцов. Для каждой выходной строки предоставляется ключ агрегации и соответствующий результат заданных операций (min, max, sum...). tAggregateRow помогает предоставить набор метрик на основе значений или Расчеты.

Основные настройки

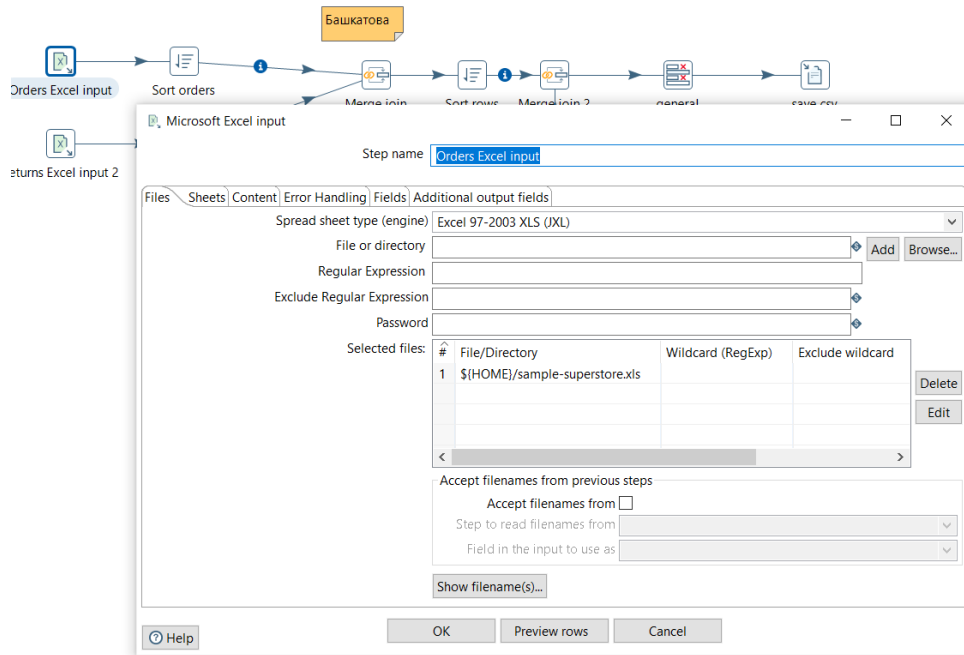
Схема и редактирование схемы	<p>Схема — это описание строки. Он определяет количество полей (столбцы), которые должны быть обработаны и переданы следующему компоненту. При создании Spark Job, избегайте зарезервированного слова при именовании Поля.line</p> <p>Нажмите «Изменить» схема, чтобы внести изменения в схему. Если текущая схема относится к типу Repository, доступны три варианта:</p> <p>Схема просмотра: выберите эту схему для просмотра только схемы.</p> <p>Изменение встроенного свойства: выберите этот параметр, чтобы изменить схему на Встроенная для локальных изменений.</p> <p>Подключение к репозиторию обновлений: Выберите этот параметр, чтобы изменить схему, хранящуюся в репозитории, и решить, следует ли распространять изменения на все задания после их завершения.</p> <p>Этот Компонент предлагает преимущество функции динамической схемы. Это позволяет: Извлечение неизвестных столбцов из исходных файлов или копирование пакетов столбцов из источника без сопоставления каждого столбца по отдельности. Дополнительные сведения о динамических схемах см. см. <i>Руководство пользователя Talend Studio</i>.</p> <p>Этот Функция динамической схемы предназначена для извлечения неизвестных столбцов стол и рекомендуется использовать только для этой цели; Не рекомендуется для Использование создания таблиц.</p>
	<p>Встроенные: вы создаете и храните схему локально для этого компонента только.</p>

Схема и редактирование схемы	<p>Схема — это описание строки. Он определяет количество полей (столбцы), которые должны быть обработаны и переданы следующему компоненту. При создании Spark Job, избегайте зарезервированного слова при именовании Поля.line</p> <p>Нажмите «Изменить» схема, чтобы внести изменения в схему. Если текущая схема относится к типу Repository, доступны три варианта:</p> <p>Схема просмотра: выберите эту схему для просмотра только схемы.</p> <p>Изменение встроенного свойства: выберите этот параметр, чтобы изменить схему на Встроенная для локальных изменений.</p> <p>Подключение к репозиторию обновлений: Выберите этот параметр, чтобы изменить схему, хранящуюся в репозитории, и решить, следует ли распространять изменения на все задания после их завершения.</p> <p>Этот Компонент предлагает преимущество функции динамической схемы. Это позволяет: Извлечение неизвестных столбцов из исходных файлов или копирование пакетов столбцов из источника без сопоставления каждого столбца по отдельности. Дополнительные сведения о динамических схемах см. см. <i>Руководство пользователя Talend Studio</i>.</p> <p>Эта Функция динамической схемы предназначена для извлечения неизвестных столбцов стол и рекомендуется использовать только для этой цели; Не рекомендуется для Использование создания таблиц.</p>
	<p>Репозиторий: Вы уже создали схему и сохранили ее в каталоге Хранилище. Вы можете повторно использовать его в различных проектах и проектах заданий.</p>
Группировка по	<p>Определите наборы агрегирования, значения которых будут равны используется для расчетов.</p>
	<p>Выходной столбец: выберите столбец в предложенном списке на основе определенной вами структуры схемы. Ты можно добавить столько выходных столбцов, сколько вы хотите, чтобы сделать более точным Агрегаты.</p> <p>Пример: выберите страну, чтобы вычислить среднее значение значений для каждая страна списка или выберите «Страна и регион», если хотите Сравните регионы одной страны с регионами другой страны.</p>
	<p>Входной столбец: сопоставление входных данных метка столбца с вашими выходными столбцами, в случае, если выходная метка Набор агрегации должен быть другим.</p>
Операционный	<p>Выберите тип операции вместе со значением для использования для расчета и поля вывода.</p>
	<p>Выходной столбец: выберите Назначение в списке.</p>
	<p>Входной столбец: выберите входные данные столбец, из которого берутся значения для агрегирования.</p>

Схема и редактирование схемы	<p>Схема — это описание строки. Он определяет количество полей (столбцы), которые должны быть обработаны и переданы следующему компоненту. При создании Spark Job, избегайте зарезервированного слова при именовании Поля.line</p> <p>Нажмите «Изменить» схема, чтобы внести изменения в схему. Если текущая схема относится к типу Repository, доступны три варианта:</p> <p>Схема просмотра: выберите эту схему для просмотра только схемы.</p> <p>Изменение встроенного свойства: выберите этот параметр, чтобы изменить схему на Встроенная для локальных изменений.</p> <p>Подключение к репозиторию обновлений: Выберите этот параметр, чтобы изменить схему, хранящуюся в репозитории, и решить, следует ли распространять изменения на все задания после их завершения.</p> <p>Этот Компонент предлагает преимущество функции динамической схемы. Это позволяет: Извлечение неизвестных столбцов из исходных файлов или копирование пакетов столбцов из источника без сопоставления каждого столбца по отдельности. Дополнительные сведения о динамических схемах см. см. <i>Руководство пользователя Talend Studio</i>.</p> <p>Эта Функция динамической схемы предназначена для извлечения неизвестных столбцов стол и рекомендуется использовать только для этой цели; Не рекомендуется для Использование создания таблиц.</p>
	<p>Игнорировать нулевые значения: выберите параметр флажки, соответствующие именам столбцов, для которых требуется значение NULL, которое следует игнорировать.</p>

Pentaho DI

1. Microsoft Excel input



Импортируются данные из Microsoft Excel. Вкладка "Файлы" определяет основные свойства файла для вывода на этом шаге.

Название шага - Название этого шага в рабочей области преобразования.

Файл или каталог - Имя файла электронной таблицы или каталога файлов, из которого вы читаете.

обычное выражение - Включает все файлы (в заданном расположении), которые соответствуют критериям, заданным этим регулярным выражением.

Исключить регулярное выражение - Исключает все файлы (в заданном расположении), которые соответствуют критериям, указанным этим регулярным выражением.

Выбранные файлы - Список файлов, которые будут использоваться на этом шаге, в соответствии с критериями, указанными в предыдущих полях.

Примите имена файлов с предыдущего шага - Если флажок установлен, будет получен список имен файлов с предыдущего шага этого преобразования. Вы также должны указать, с какого шага вы импортируете, и поле ввода на этом шаге, из которого вы будете извлекать данные имени файла. Если вы выберете этот параметр, параметр Показывать имена файлов будет показывать предварительный просмотр списка имен файлов.

Вкладка "Листы" - Вкладка "Листы" указывает, какие листы вы хотите использовать в указанных файлах. Документ электронной таблицы может содержать несколько листов.

2. Sort rows

Башкатова

ers Excel input

Sort orders

Sort rows

Merge join

Sort rows

Merge join 2

general

save csv

ns Excel input 2

Sort rows

Step name: Sort orders

Sort directory: %%java.io.tmpdir%%

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files?: ☐

Only pass unique rows? (verifies keys only): ☐

Fields:

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	Order ID	Y	N	N	0	N

Help OK Cancel Get Fields

Сортировка строк во входных потоках на основе определенных критериев.

3. Merge join

Башкатова

orders

Sort rows

Merge join

Sort rows

Merge join 2

general

returns

Merge join

Step name: Merge join

First Step: Sort orders

Second Step: Sort returns

Join Type: LEFT OUTER

Keys for 1st step:

#	Key field
1	Order ID

Get key fields

Keys for 2nd step:

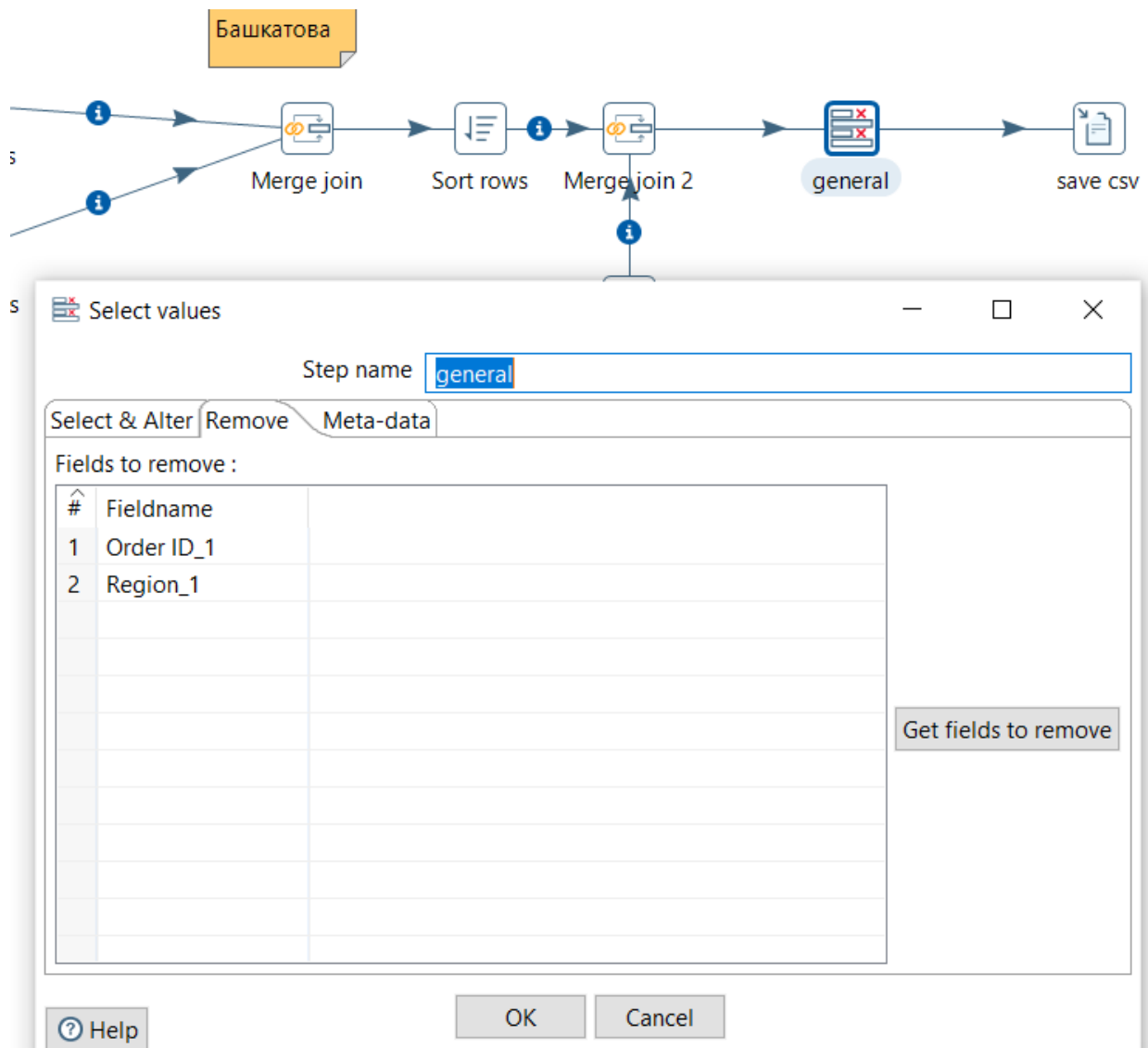
#	Key field
1	Order ID

Get key fields

Help OK Cancel

Шаг объединения слиянием выполняет ЛЕВОЕ ВНЕШНЕЕ соединение для двух наборов. В результате получается список. "Объединение слиянием" требует, чтобы входные данные для шага были отсортированы.

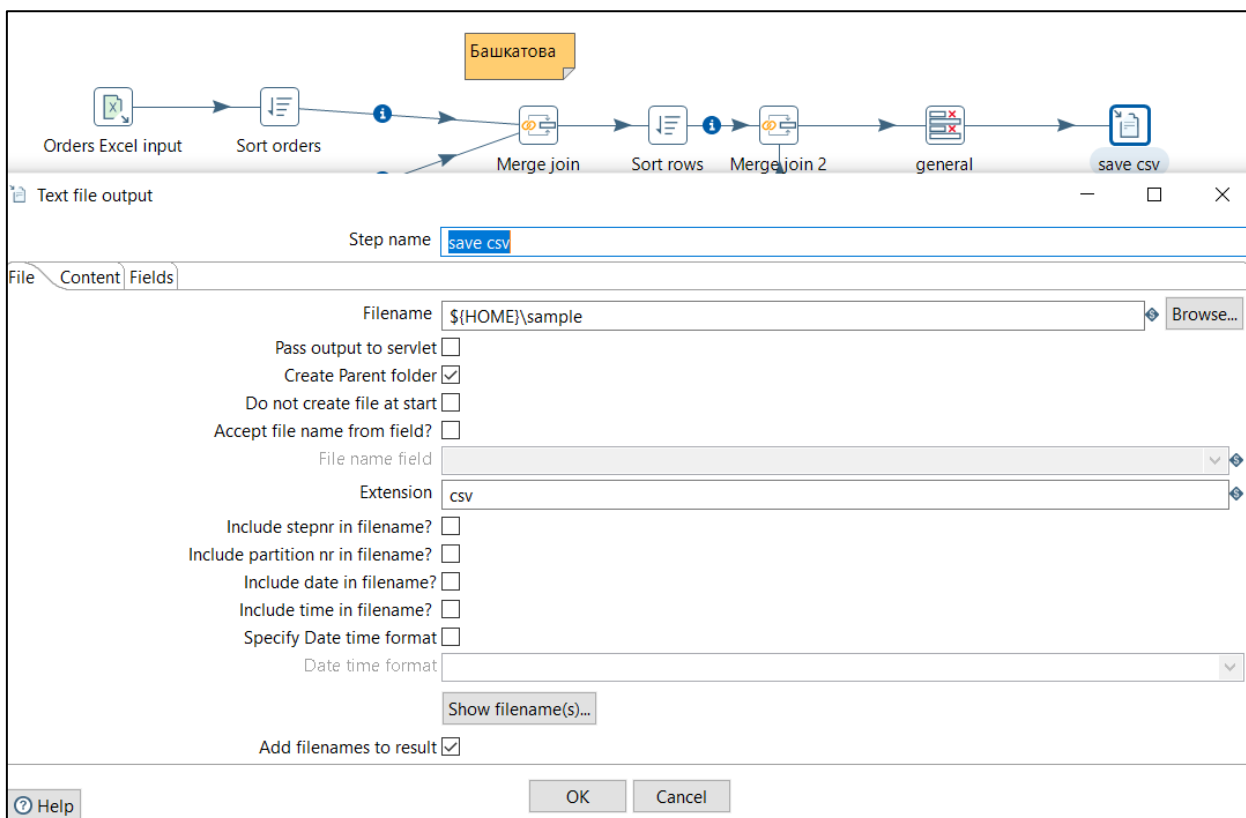
4. Select values



Шаг выбора значений полезен для выбора, удаления, переименования, изменения типов данных и настройки длины и точности полей в потоке. Эти операции разбиты на различные категории:

- Выбрать и изменить - укажите точный порядок и название, в котором поля должны быть размещены в выходных строках.
- Удалить - укажите поля, которые должны быть удалены из выходных строк.
- Метаданные - измените имя, тип, длину и точность (метаданные) одного или нескольких полей.

5. Text file output



Шаг вывода в текстовый файл используется для экспорта данных в формат текстового файла. Это обычно используется для создания запятой. Разделенные значения (CSV-файлы), которые могут быть прочитаны приложениями для работы с электронными таблицами. Также возможно создавать файлы фиксированной ширины, задав длину полей на вкладке поля.

Название шага-При желании вы можете изменить название этого шага в соответствии с вашими потребностями.

Имя файла-Укажите CSV-файл, из которого следует записывать

Вместо этого запустите это как команду?-Включите возможность "передавать" результаты в указанную вами команду или скрипт

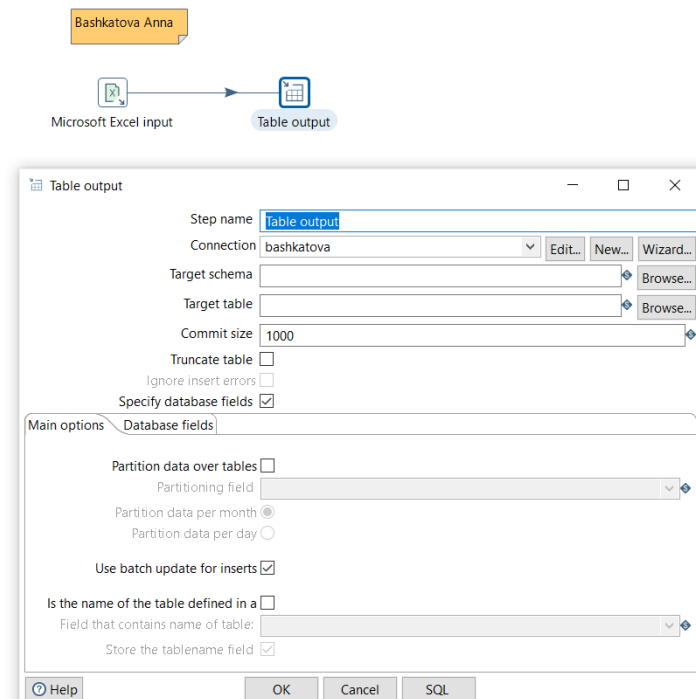
Принять имя файла из поля?-Разрешить указывать имя (имена) файла(ов) в поле во входном потоке

Поле имени файла-Если включена предыдущая опция, вы можете указать поле, которое будет содержать имя файла(ов) во время выполнения.

Расширение-Добавляет точку и расширение в конец имени файла. (.txt)

Включите stepnr в имя файла-Если вы выполняете шаг в нескольких копиях (запускаете несколько копий шага), номер копии указывается в имени файла перед расширением. (_0).

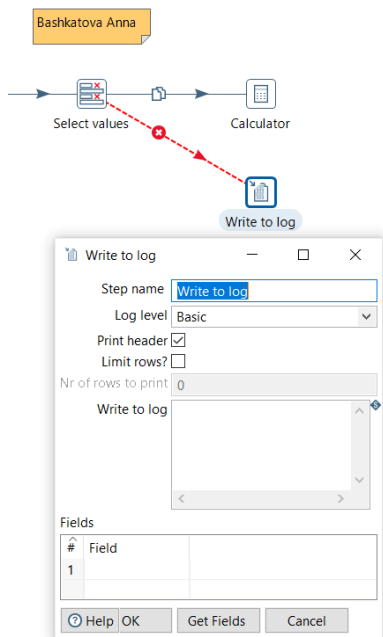
6. Table output



Шаг вывода таблицы позволяет загрузить данные в таблицу базы данных. Вывод таблицы эквивалентен оператору DML, INSERT. Этот шаг предоставляет параметры конфигурации для целевой таблицы и множество параметров, связанных с обслуживанием и / или производительностью, таких как размер фиксации и использование пакетного обновления для вставок.

Если у вас есть база данных Postgres или MySQL, в которой есть столбцы идентификаторов, и вы вставляете запись как часть вставки, драйвер JDBC обычно возвращает автоматически сгенерированный ключ, который он использовал при выполнении вставки.

7. Write to log



Запись данных в журнал.

8. Number range

Bashkatova Anna

Data grid

Calculator

1.3.
2.5

performance

message

1.3.
2.5

Number range

Step name:performance

Input field:diff_dates

Output field:range

Default value(if no range)unknown

Ranges (min <= x < max):

#	Lower Bound	Upper Bound	Value
1		30.0	excellent
2	30.0	180.0	very good
3	180.0	360.0	good
4	360.0		poor

Help

OK

Cancel

Создавайте диапазоны на основе числовых полей.

Вариант	Определение
Название шага	Название этого шага в том виде, в каком оно отображается в рабочей области преобразования .
Поле ввода	Укажите поле для создания диапазонов.
Выходное поле	Укажите название поля диапазонов, которое будет создано.
Значение по умолчанию	Значение, возвращаемое, если в указанных диапазонах нет совпадений .
Диапазоны	Обозначил верхнюю и нижнюю границы диапазона.
Нижняя граница	Обозначает минимальное значение диапазона.
Верхняя граница	Укажите верхнее значение диапазона.
Ценность	Назначил имя для значения.