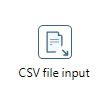
Выполнила: Вереина Мария, АДЭУ-211

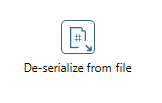
Практическая работа №3

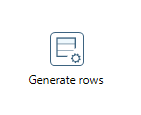
Подсистемы Pentaho DI:

1. Подсистема Input. Компоненты:

**CSV file input.** Этот шаг предоставляет возможность считывать данные из файла с разделителями. Можно определить любой разделитель, который хотите использовать, например, табуляции и точки с запятой; вы не ограничены использованием запятых. Внутренняя обработка позволяет на этом этапе быстро обрабатывать данные.

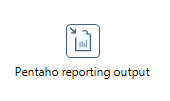
**Data grid.** Этот шаг позволяет ввести статический список строк в таблице. Обычно это делается для тестирования, справочных или демонстрационных целей.

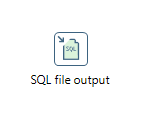
**The De-serialize from file** считывает строки данных из двоичного файла Kettle, содержащего строки и метаданные. Рекомендуется использовать этот шаг только для хранения данных с коротким сроком службы, Pentaho не может гарантировать, что формат файла останется неизменным в разных версиях Pentaho Data Integration.

**Generate rows** выводит указанное количество строк. По умолчанию строки пустые; однако они могут содержать несколько статических полей. Этот шаг используется в основном для целей тестирования. Это может быть полезно для генерации фиксированного количества строк. Иногда с Generate Rows создают одну строку, которая является отправной точкой для вашего преобразования.

Шаг **Google Analytics** в разделе "Ввод", позволяет получить доступ к данным Google Analytics для создания отчетов или заполнения хранилища данных. Он использует API Google Analytics 3.0. Google Analytics Query Explorer предоставляет специальный веб-сайт для разработки и тестирования запросов с помощью API Google Analytics, независимо от преобразования PDI.

2. Подсистема Output. Компоненты:

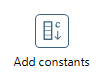
**Pentaho** **reporting** **output**. На этом шаге отображается отчет, разработанный Pentaho Report Designer, так называемый prpt-файл. Доступны различные параметры вывода, доступные для механизма создания отчетов Pentaho (PDF, HTML, Excel, ...).

**SQL file output.** На этом шаге входные данные записываются в текстовый файл в виде набора инструкций SQL. SQL генерируется на диалекте выбранной базы данных.

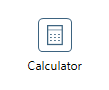
**Update** аналогично шагу Update Input, за исключением того, что выполняются только обновления. Вставки не выполняются.

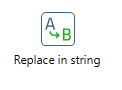
3. Подсистема Streaming. Компоненты:

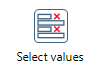
**The Java Messaging Service** (JMS) позволяет Pentaho DI получать текстовые сообщения с сервера Apache ActiveMQ JMS. Можно получать текстовые сообщения с любого JMS-сервера с поддержкой JNDI, если встроить PDI для запуска в контейнере J2EE. Вы могли бы использовать JMS Consumer для определения долговременного преобразования, которое обновляет хранилище данных каждый раз при получении сообщения JMS.

4. Подсистема Transform. Компоненты:

**Add constants** - это простой и высокопроизводительный способ добавления постоянных значений в поток.

**Calculator**. Этот шаг предоставляет вам предопределенные функции, которые могут быть выполнены для значений полей ввода.

**Replace** **in** **string** — это простой поиск и замена. Он также поддерживает регулярные выражения и ссылки на группы. ссылки на группы выбираются в строке replace by как $n, где n - номер группы.



**Select Values** полезен для выбора, удаления, переименования, изменения типов данных и настройки длины и точности полей в потоке. Эти операции организованы в разные категории:

Выберите и измените — укажите точный порядок и имя, в котором поля должны быть размещены в выходных строках

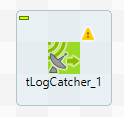
Удалить — укажите поля, которые должны быть удалены из выходных строк

Метаданные - изменение имени, типа, длины и точности одного или нескольких полей

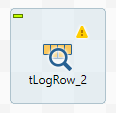
**Sort rows** сортирует строки на основе заданных вами полей и того, следует ли их сортировать в порядке возрастания или убывания.

Talend Components

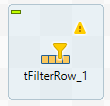
5. Журналы и ошибки. Компоненты:

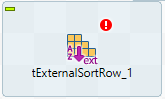
 Извлекает установленные поля и сообщения из Java Exception, die и/или warn и передает их следующему компоненту.

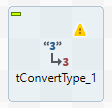
 Этот компонент выдает ошибку и завершает работу. Если вы просто хотите выдать предупреждение, обратитесь к документации по предупреждению.

 Отображает данные или результаты в консоли запуска для мониторинга обработанных данных.

6. Обработка. Компоненты:

 Фильтрует входные строки, устанавливая одно или несколько условий для выбранных столбцов.

 Сортирует входные данные на основе одного или нескольких столбцов, по типу сортировки и порядку, используя внешнее приложение сортировки.

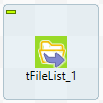
 Автоматически преобразует один тип Talend java в другой и, таким образом, позволяет избежать ошибок компиляции.

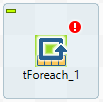
tConvertType позволяет выполнять определенные преобразования во время выполнения из одного типа Talend java в другой.

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

tMap — это продвинутый компонент, который интегрируется в качестве плагина в Talend Studio.

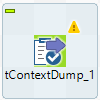
7. Подсистема Оркестровка. Компоненты:

 Выполняет итерацию набора файлов или папок в заданном каталоге на основе шаблона маски файла. Этот компонент выполняет итерацию по каждому файлу в каталоге, включая системный файл, скрытый файл, файл с нулевым байтом и так далее, до тех пор, пока файл соответствует условиям, заданным в поле Файлы.

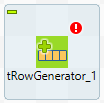
 Создает цикл в списке для повторяющейся ссылки.

 Выполняет задачу или поручение автоматически, на основе цикла

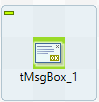
8. Подсистема Разное. Компоненты:

 Копирует настройки контекста текущего задания в плоский файл, таблицу базы данных и т.д., Которые затем могут быть использованы tContextLoad.

Вместе с tContextLoad этот компонент упрощает применение настройки контекста одного задания к другому.

 Создает входной поток в задании для целей тестирования, в частности, для граничных наборов тестов.

tRowGenerator генерирует столько строк и полей, сколько требуется, используя случайные значения, взятые из списка.

 Открывает диалоговое окно с кнопкой OK, требующей действий от пользователя.

msgbox - это графический перерыв в ходе выполнения задания.