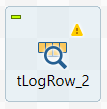
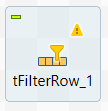
Выполнила: Белик Мария,

АДЭУ-211

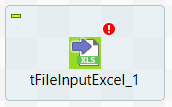
Практика №3

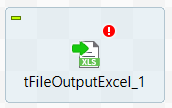
Компоненты Talend:

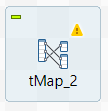
– **отслеживает** обработку данных. Он отображает данные / вывод в консоли запуска.

 - tFilterRow будут фильтровать или ограничивать входные строки на основе заданного выражения и возвращать выходные данные. Мы можем использовать это поле Talend Filter Row, чтобы получить продажи выше определенного значения или вернуть записи определенного города, штата или страны и т. д.

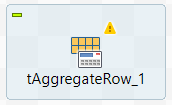
 - tJoin объединяет две таблицы, выполняя точное совпадение в нескольких столбцах. Он сравнивает столбцы из основного потока со справочными столбцами из потока поиска и выводит данные основного потока и/или отклоненные данные.

 - Читает файл Excel построчно, чтобы разбить их на поля с помощью регулярных выражений, а затем отправляет поля, как определено в схеме, в следующий компонент.

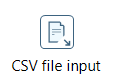
 - Записывает файл MS Excel с разделенными значениями данных в соответствии с определенной схемой.

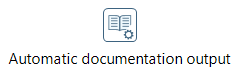
 - Преобразует и направляет данные из одного или нескольких источников в одно или несколько мест назначения. tMap — это расширенный компонент, который интегрируется в качестве плагина в Talend Studio.

- Компонент **tAssert** является частью семейства компонентов [Logs & Errors](https://www.talendbyexample.com/talend-logs-and-errors-component-reference.html) и отправляет *неблокирующее* сообщение в [tAssertCatcher](https://www.talendbyexample.com/talend-tassertcatcher-component.html) .

 - Получает поток и агрегирует его на основе одного или нескольких столбцов. Для каждой строки вывода предоставляется ключ агрегации и соответствующий результат операций набора (минимум, максимум, сумма...). tAggregateRow помогает предоставить набор показателей на основе значений или вычислений.

Компоненты Pentaho DI:

- Шаг ввода CSV используется для чтения данных из файлов с разделителями. Это подмножество шага ввода текстового файла. Имя CSV для этого шага неверно, и этот шаг можно использовать для чтения данных из любого файла с разделителями.

- Этот шаг используется для создания описательной документации для одного или нескольких преобразований или заданий. Это можно использовать как способ автоматического создания документации о назначении заданий и преобразований или как способ архивирования их поведения по мере их изменения с течением времени.

- Этот шаг предоставляет вам предопределенные функции, которые могут быть выполнены над значениями полей ввода.

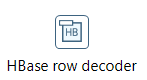
Помимо аргументов (поле A, поле B и поле C) необходимо также указать возвращаемый тип функции. Вы также можете

удалить поле из результата (вывода) после вычисления всех значений.

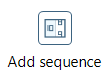
- Шаг Avro Input декодирует двоичные или JSON данные Avro и извлекает поля из структуры, которую он определяет, либо из плоских

файлов или из входящих полей.

- Шаг преобразования Unique Rows (HashSet) отслеживает точные повторяющиеся строки. Этот шаг также может удалить повторяющиеся строки и оставить только уникальные вхождения.  шаг «Unique Rows (HashSet)» не требует отсортированных входных данных для обработки повторяющихся строк, вместо этого он отслеживает дубликаты в памяти.

- Шаг HBase Row Decoder декодирует входящий ключ и объект результата HBase в соответствии с сопоставлением.

- Шаг Concat Fields используется для объединения нескольких полей в одно целевое поле. Поля могут быть разделены разделителем, а логика включения полностью совместима с этапом вывода текстового файла.

- Шаг добавляет последовательность в поток. Последовательность — это постоянно изменяющееся целочисленное значение с определенным начальным значением и значением приращения. Вы можете либо использовать последовательность базы данных для определения значения последовательности, либо сгенерировать ее с помощью Kettle.