**Pentaho**

Всего 4 группы:

1) Data Extracting – получение данных

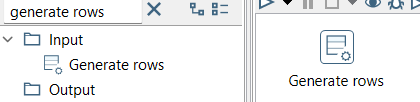
2) Cleaning and Conforming Data – интеграция данных и подготовка к загрузке

3) Delivering Data for Presentation - обработка данных в DW

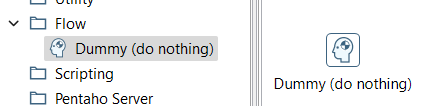
4) Managing the ETL environment – управление и мониторинг компонентами ETL

Компоненты:

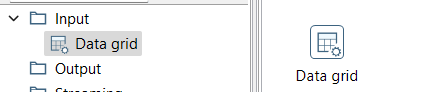
1) Generate row – создание строк



2) Dummy - компонент для завершения какой - либо операции. Пример: завершения операции Generate row



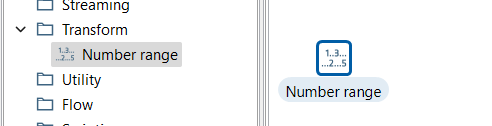
3) Data grid – создает статистические значения



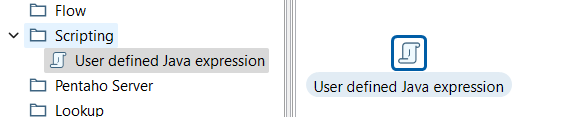
4) Calculator – проводит математические операции с полями данных



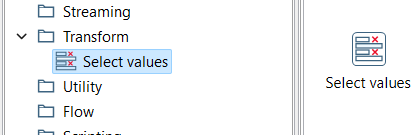
5) Number range – создает последовательность чисел.



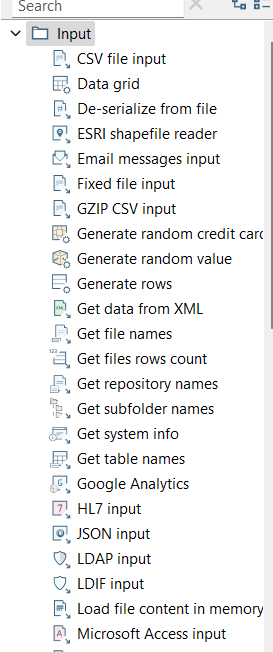
6) User defined Java expression - для выполнения функций используется Java код



7) Select values – проводит изменения в поле в строке. Пример: изменения типа данных

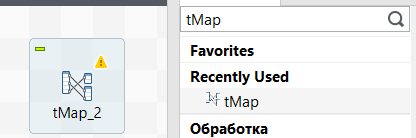


8) Input – загружает данные из выбранной категории

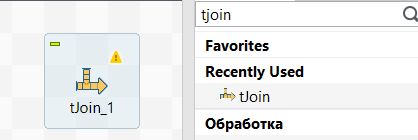


**Talend:**

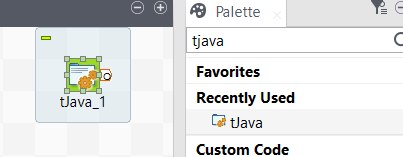
1) **tMap** – Берет данные из одного или нескольких источников, преобразует их, а затем отправляет преобразованные данные одному или нескольким получателям.



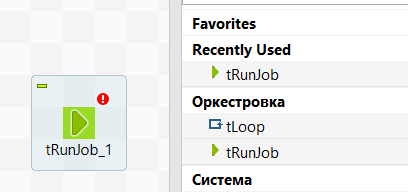
2) **tJoin** – объединяет 2 таблицы, выполняя внутренние и внешние соединения между основным потоком и потоком поиска.



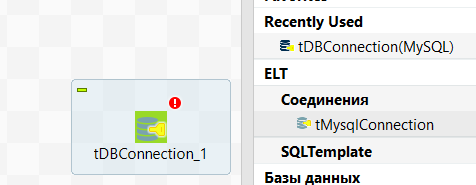
3) **tJava** – позволяет использовать персонализированный код Java в программе Talend.



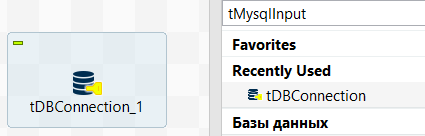
4) **tRunJob** – Управляет сложными системами заданий, выполняя одно задание Talend за другим.



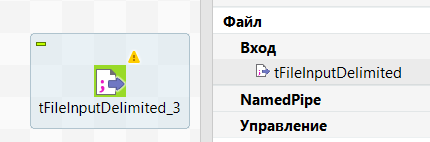
5) **tMysqlConnection** – **подключается** к базе данных MySQL, определенной в компоненте.



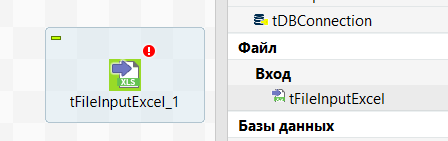
6) **tMysqlInput** – запускает запрос к базе данных, чтобы прочитать базу данных и извлечь поля (таблицы, представления и т. д.) в зависимости от запроса.



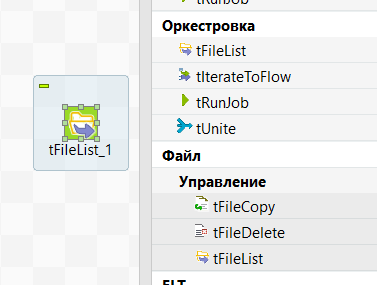
7) **tFileInputDelimited** – Считывает файл за строкой с разделителями, разделяет их на отдельные поля и передает их следующему компоненту.



8) **tFileInputExcel** – Считывает файл Excel по строкам, разделяет их на отдельные поля и передает их следующему компоненту.



9) **tFileList** – получает все файлы и каталоги из заданного шаблона маски файла.



10) **tFileArchive** – сжимает набор файлов или папок в архивный файл zip, gzip или tar.gz.

