








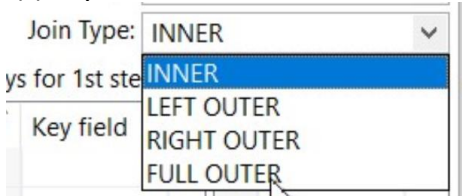




## Основные компоненты ETL-подсистем в PENTANO BI

Кузьменко Сергей БД-231. Лабораторная 4.1

Компонент	Описание
<p>Table input</p>  <p>stg.orders</p>	<p>Считывание данных из таблицы в базе данных. Необходимо настраивать подключение к базе</p>
<p>Excel input и Text file input</p>   <p>Microsoft Excel inputText file input</p>	<p>Считывание информации из файла разных форматов. Необходимо указывать путь к файлу и файловой системе</p>
<p>Select values</p>  <p>Select values</p>	<p>Отбор полей или удаление полей в исходной таблице</p>
<p>Add sequence</p>  <p>Add sequence 2</p>	<p>Генерирование нумерации строк (последовательности чисел). Можно использовать для создания первичного ключа</p>
<p>Calculator</p>  <p>Calculator</p>	<p>Используется для вычислений в таблице. Доступны множества функций</p> <p>filter: [ ]</p> <p>Select the calculation type to perform</p> <ul style="list-style-type: none"> <li>-</li> <li>Set field to constant value A</li> <li>Create a copy of field A</li> <li>A + B</li> <li>A - B</li> <li>A * B</li> <li>A / B</li> <li>A * A</li> <li>SQRT( A )</li> <li>100 * A / B</li> <li>A - ( A * B / 100 )</li> <li>A + ( A * B / 100 )</li> <li>A + B * C</li> <li>SQRT( A*A + B*B )</li> <li>ROUND( A )</li> <li>ROUND( A , B )</li> <li>STDROUND( A )</li> <li>STDROUND( A , B )</li> <li>CEIL( A )</li> <li>FLOOR( A )</li> <li>NVL( A , B )</li> <li>Date A + B Days</li> <li>Year of date A</li> <li>Month of date A</li> <li>Day of year of date A</li> <li>Day of month of date A</li> <li>Day of week of date A</li> <li>Week of year of date A</li> <li>ISO8601 Week of year of date A</li> <li>ISO8601 Year of date A</li> <li>Byte to hex encode of string A</li> </ul>

<p>Unique rows</p>  <p>Unique rows 2</p>	<p>Оставляет только уникальные строки по выбранному столбцу, группе столбцов или по всем столбцам</p>
<p>Group by</p>  <p>Group by</p>	<p>Позволяет группировать данные и использовать агрегаты</p>
<p>Merge join</p>  <p>Merge join</p>	<p>Слияние таблиц. Таблицы необходимо отсортировать по ключу соединения перед Merge, иначе функция работает неверно. Об этом предупреждает Pentaho. Доступны все типы соединения</p> 
<p>Table output</p>  <p>Table output</p>	<p>Загрузка данных в таблицы БД. Необходимо настраивать подключение к БД. Также важно ставить флаг</p>  <p>Чтобы данные перезаписывались при каждом запуске</p>