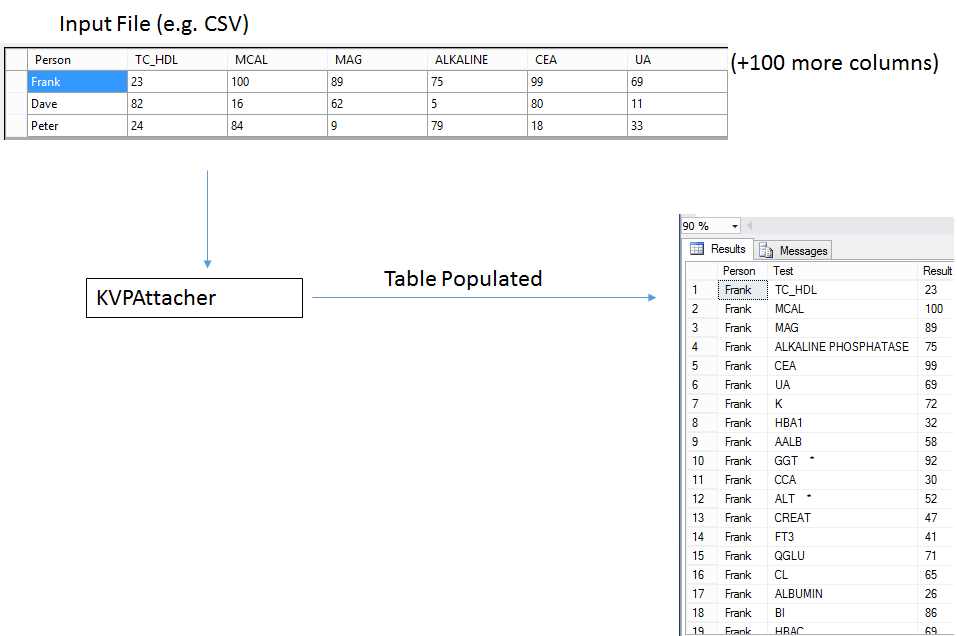
KVP Attacher

# Background

KvpAttacher is a component for use in a data load (See UserManual.docx). As an Attacher it is responsible for populating the RAW data load bubble with records. It is designed to take an input file with a variable number of columns (usually a large - e.g. 8000 columns) and load a table in the RAW database with Key/Value pair equivalents. This results in a manageable number of stable columns in the table schema.



In the above example we can see an input file with many columns. The first column is special in that it identifies who the record belongs to followed by lots and lots of measurements for that person. We want to split each column into a Key/Value pair since this lets us have a sane data model (see Table Populated above). But we also want to maintain the ‘Person’ in every row so that we know who the test/result belongs to.

Within KVPAttacher the common column(s) are referred to as Primary Key Column(s) because they are unique in the input file and identify which source record a given Key/Value pair match. In the above example the column ‘Person’ is the only Primary Key column.

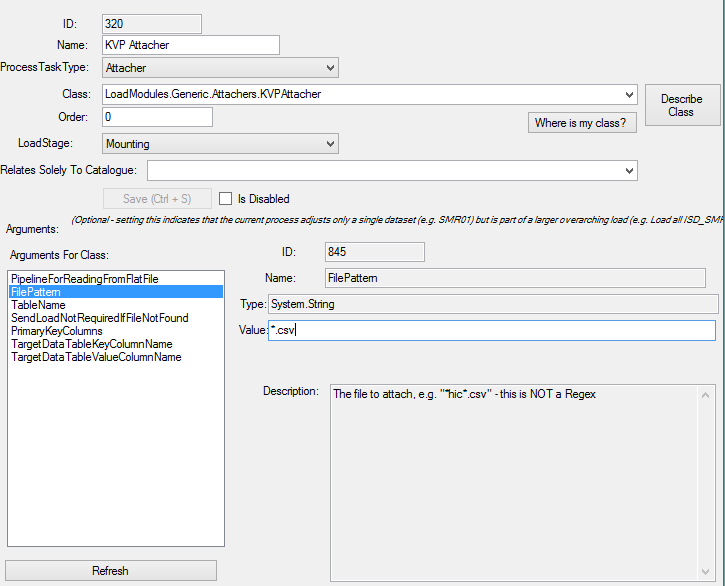
All other columns in the input table are converted into key/value pairs with the ‘key’ being the column header and the ‘value’ being the value in the cell. You must also specify the TargetColumns for the Key (‘Test’ in the above example) and Value (‘Result’ in the above example).

# Configuring KVPAttacher

## Basic Settings

In order to begin using KVPAttacher you must already have a schema table (which might be empty) in your live database and have imported it into the RMDP as a Catalogue. You must configure a new data load to target the dataset (See UserManual.docx).

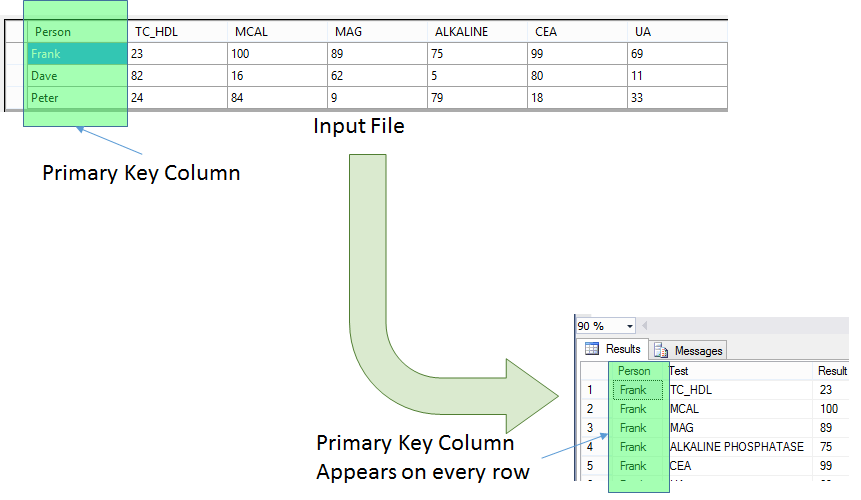
Place the files you want to load into the forLoading data load folder. Then add a new ProcessTask at the Mounting stage and set its Class to KVPAttacher.

 Start by configuring the TableName which is the name of the table in RAW that will be populated by the attacher (same as the name of your live table).

Next configure the FilePattern such that it matches your input file(s) e.g. \*.csv

## Key, Value & Primary Key Settings

Primary Key Column(s) are optional. They refer to columns that appear in the data table which are not reinterpreted into KeyValue pairs. Instead each column is preserved into the key / value row with its original value.



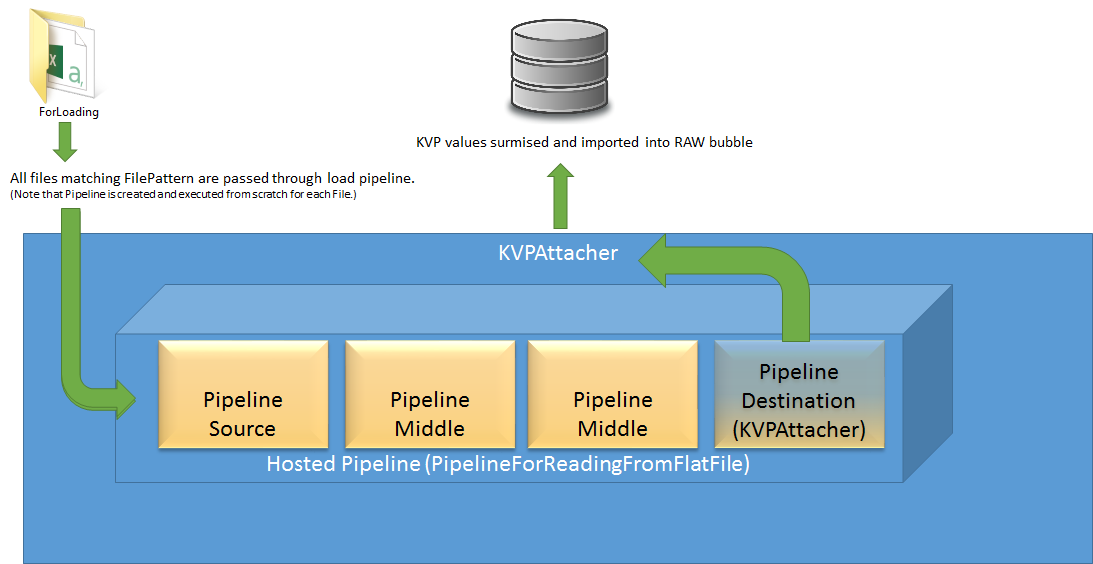
TargetDataTableKeyColumnName is the name of the column in your live table which will be populated with the column headers (keys). TargetDataTableValueColumnName is the name of the column in your live table which will be populated with the corresponding cell values (values).

## Configuring Pipeline Setting

KVPAttacher needs to open and read from an input file (e.g. csv file) rather than duplicate the functionality of opening each file type (and in order to maximise versatility), KVPAttacher hosts a Pipeline which it expects to handle all the file reading/adjusting for it.

This means that it can handle any input file type that you have a compatible pipeline source for (e.g. tab separated, comma separated, excel etc). It also means you can add adjustment components (e.g. Transposer) to make changes to the Flat File as it is loaded into memory.

KVPAttacher requires that the Pipeline you select be a flow of DataTable (obviously) and will pass a single input object (FlatFileToLoad).



In practice this is a lot simpler than it sounds. Start by clicking on the PipelineForReadingFromFlatFile argument and select ‘Creating New…’ to create a new Pipeline. You should see a fixed destination (KVPAttacher) and an empty slot for a source. Drag either a DelimitedFlatFileDataFlowSource (for csv, tab delimited etc) or an ExcelDataFlowSource.

