Changing Data Layouts

Why Pivot?

There are many reasons why you may want to change the layout of your data. Perhaps you need to use the column headers as data values, or you need the values in your dataset to be aligned in a particular way in order to use a certain tool. Pivoting data can also give you a new perspective and spark new ideas. Whatever your reason for wanting to change the layout of your data, the Transpose and Cross Tab tools are flexible options to help you get your data looking the way you want.

!=Transpose()

If you are familiar with the Transpose function in excel, you may be asking why there are two tools for pivoting the layout of data. An important factor to keep in mind is that unlike excel, you cannot specify an array in Designer when applying a function. Rather, changes are applied to an entire dataset. Additionally, Designer distinguishes between rows and columns. Some tools need data to be arranged in a specific way (e.g. formula vs summarize). Similarly, the Transpose tool pivots horizontal data to vertical orientations, while the Cross Tab tool pivots vertical data to horizontal orientations.

To use these tools well, you must have a clear idea of the way you want your dataset to look when you are finished. If you are not sure, the configuration options can become challenging, especially when working with larger datasets. Once you know what you want to do, a basic understanding of the configuration options for each tool is sufficient for you to transform your data in powerful ways.

The Transpose Tool

When transposing data, the tool will output at least 2 columns of data which have standardized names. The column called Name contains the Headers from the incoming dataset. The column called Value contains the values from your dataset, aligned vertically. These columns are created in tandem. The first row under Name is populated by the header from the first column of the incoming dataset. The first row under Value is populated by the value from the first row of the first column of the incoming dataset. The second row under name contains the header from the second column of the incoming dataset. The first value under the second column of the incoming dataset populates the second row for Value. When Designer has gotten to the last column from the incoming dataset, the pattern is repeated in the Name column, but the Values will be populated by the values in the next row of the incoming dataset. When finished, the number of records will equal the number of incoming columns times the number of rows in the incoming dataset.

The Transpose tool's configuration options allow you to select what data is transposed using the Data Columns checkboxes. Designer will populate this list with the columns of

the incoming dataset. By checking a value, it will be included in the Name and Value outputs.

Key Columns
The Transpose tool's configuration window also includes a-s- section called "Key Columns". When you select a column in this section, it will be excluded from the transpose process and removed from the data columns section automatically. By selecting a key column, it will maintain its current orientation (header and data values). Since the number of values in the key column is usually less than the number of values expected in the Name and Value outputs, values in the key column are replicated until the pattern restarts.

Multiple key columns can be selected. Each key column will add one column to the output, in addition to the 2 default columns (name and Value). If a column is not selected as a key column or a data column, it is dropped from the data stream.

Given the pattern of the output, drag the highlighted cell into its position in the incoming dataset.

The Cross Tab Tool
Using the Cross Tab tool to pivot vertical data to a horizontal layout can be trickier. While the Transpose tool creates a pattern when pivoting data, the Cross Tab tool requires a pattern in order to pivot. In most instances, the goal is not to simply turn vertical data into a horizontal layout. Rather, it is a more common goal to sort data into appropriate columns and rows. Since the data will be placed into a smaller number of columns than there are rows of data, it makes sense that the associated columns (destination) would repeat. Unfortunately, most incoming datasets will not have a readymade pattern that corresponds to the columns desired in the output. In those instances, it is up to the user to create or condition a column that will serve as the pattern before using the Cross Tab tool. If you don't have a clear idea of what the output should look like, this will be very challenging indeed.

To use the Cross Tab tool, at least 2 columns must be present in the incoming dataset. One will serve as the column headers in the output, and the other will serve as the data values in the output. In the Cross Tab tool's configuration window, dropdowns allow you to select an incoming column to use as column headers, and a column for populating values in the output. Notice that only one column can be specified for values. This means that all data that should be present in the output must be in one column when coming into the Cross Tab tool.

The incoming values used as columns will be aggregated. In the same way that the Transpose tool replicated column headers when creating the "Name" column, the Cross Tab does the opposite action; removing duplicate values. This means that when you only

specify two incoming columns (headers and values), only one row will be created. *record scratch sound* When a header value occurs multiple times, that indicates that multiple values are associated with that output column. Since the header will only appear once, multiple rows are listed under a column. However, Designer will not make assumptions about which values belong on a particular row, after all, values on a row can also be associated with each other. With only two incoming columns (aka pieces of information), Designer cannot guess which row a value should populate. As such, one row is created and a "method for aggregating values" is required.

Aggregating Values
In the Cross Tab tool's configuration window, the "Method for Aggregating Values" section provides a list of available options for combining values associated with a given header value. These options are determined by the datatype of the incoming column specified in the "Values for New Columns" dropdown.

Columns with a String datatype will have options to concatenate the values, or use the first or last value associated with the header.

Columns with a Numeric datatype will have options to Sum, Average, Count, use first or last, and several other options.

Grouping in Cross Tab
But what if you do want more than one row to appear in the output. In that case, a third column will be used to group data and provide the missing row information. Like the Header values, these values will be consolidated and should form a pattern matching the rows desired in the output. It can be helpful to think of the header and grouping values as X and Y coordinates. The column header would be the X value and the grouping value represents the Y coordinate. Designer will place a value in the cell specified by the values in the header and grouping columns.

In most cases the Cross Tab tool will require that other tools "condition" the data before it can be pivoted. Any incoming values that are not selected in the Cross Tab tool's configuration are dropped from the data stream. If those values are important, a join may be required. There are also occasions where the Transpose and Cross Tab tools will need to be used multiple times in the same stream. The tools used will depend on the incoming dataset and the type of pivot desired, but it is always important to know what the data should look like in the end state.

Best Practices
Other things to keep in mind when using the Transpose and Cross Tab tools:

- If the incoming dataset includes multiple datatypes when transposing data columns, the values column will automatically convert to the most robust datatype to accommodate the various values. (typically VString)

- The Cross Tab tool will replace special characters and spaces in headers with underscores to increase speed and efficiency when processing. After cross tabbing, you may find you need to re-name your columns to remove the underscores.

- The Cross Tab tool automatically sorts headers and groups in ascending order. This is an important consideration when joining with other datastreams, especially by position.