Wrangling Data Flow Functions

Contents

[Wrangling Data Flow 1](#_Toc7348640)

[Column Management 1](#_Toc7348641)

[Row Filters 1](#_Toc7348642)

[Adding and Transforming Columns 1](#_Toc7348643)

[Merging/Joining tables 2](#_Toc7348644)

[Group by 2](#_Toc7348645)

[Sorting 2](#_Toc7348646)

[Notable Unsupported Functionality (not exhaustive; subject to change) 2](#_Toc7348647)

Wrangling Data Flow

Wrangling Data Flow in Azure Data Factory integrates with [Power Query Online](https://docs.microsoft.com/en-us/powerquery-m/power-query-m-reference) and makes the best in class Power Query M functions available for data wrangling @ cloud scale via spark execution.

Column Management:

Selection, Removal, Renaming, and Reordering (corresponding M functions: [Table.SelectColumns](https://docs.microsoft.com/en-us/powerquery-m/table-selectcolumns), [Table.RemoveColumns](https://docs.microsoft.com/en-us/powerquery-m/table-removecolumns), [Table.RenameColumns](https://docs.microsoft.com/en-us/powerquery-m/table-renamecolumns), [Table.ReorderColumns](https://docs.microsoft.com/en-us/powerquery-m/table-reordercolumns), [Table.PrefixColumns](https://docs.microsoft.com/en-us/powerquery-m/table-prefixcolumns), [Table.TransformColumnNames](https://docs.microsoft.com/en-us/powerquery-m/table-transformcolumnnames))

## Row Filters

(corresponding M function: [Table.SelectRows](https://docs.microsoft.com/en-us/powerquery-m/table-selectrows)):

* Numeric, Text, and Logical Equality/Inequality
* Numeric comparisons
* Text containment (corresponding M functions: [Text.Contains](https://docs.microsoft.com/en-us/powerquery-m/text-contains), [Text.StartsWith](https://docs.microsoft.com/en-us/powerquery-m/text-startswith), [Text.EndsWith](https://docs.microsoft.com/en-us/powerquery-m/text-endswith))
* Combinations of these (and/or/not)

## Adding and Transforming Columns

(corresponding M functions: [Table.AddColumn](https://docs.microsoft.com/en-us/powerquery-m/table-addcolumn), [Table.TransformColumns](https://docs.microsoft.com/en-us/powerquery-m/table-transformcolumns), [Table.ReplaceValue](https://docs.microsoft.com/en-us/powerquery-m/table-replacevalue), [Table.DuplicateColumn](https://docs.microsoft.com/en-us/powerquery-m/table-duplicatecolumn)):

* Numeric Arithmetic
* Text concatenation
* Most standard, scientific, and trigonometric numeric functions (corresponding M functions: [Number.Abs](https://docs.microsoft.com/en-us/powerquery-m/number-abs), [Number.Log](https://docs.microsoft.com/en-us/powerquery-m/number-log), [Number.Log10](https://docs.microsoft.com/en-us/powerquery-m/number-log10), [Number.Mod](https://docs.microsoft.com/en-us/powerquery-m/number-mod), [Number.Power](https://docs.microsoft.com/en-us/powerquery-m/number-power), [Number.Sqrt](https://docs.microsoft.com/en-us/powerquery-m/number-sqrt), and all functions under [Trigonometry](https://docs.microsoft.com/en-us/powerquery-m/number-functions#trigonometry))
* Replacement (corresponding M functions: [Replacer.ReplaceText](https://docs.microsoft.com/en-us/powerquery-m/replacer-replacetext), [Replacer.ReplaceValue](https://docs.microsoft.com/en-us/powerquery-m/replacer-replacevalue), [Text.Replace](https://docs.microsoft.com/en-us/powerquery-m/text-replace))
* Positional Text Extraction (corresponding M functions: [Text.PositionOf](https://docs.microsoft.com/en-us/powerquery-m/text-positionof), [Text.Length](https://docs.microsoft.com/en-us/powerquery-m/text-length), [Text.Start](https://docs.microsoft.com/en-us/powerquery-m/text-start), [Text.End](https://docs.microsoft.com/en-us/powerquery-m/text-end), [Text.Middle](https://docs.microsoft.com/en-us/powerquery-m/text-middle), [Text.RemoveRange](https://docs.microsoft.com/en-us/powerquery-m/text-removerange))
* Basic Text Formatting (corresponding M functions: [Text.Lower](https://docs.microsoft.com/en-us/powerquery-m/text-lower), [Text.Upper](https://docs.microsoft.com/en-us/powerquery-m/text-upper), [Text.Trim](https://docs.microsoft.com/en-us/powerquery-m/text-trim)/[Start](https://docs.microsoft.com/en-us/powerquery-m/text-trimstart)/[End](https://docs.microsoft.com/en-us/powerquery-m/text-trimend), [Text.PadStart](https://docs.microsoft.com/en-us/powerquery-m/text-padstart)/[End](https://docs.microsoft.com/en-us/powerquery-m/text-padend))
* If expressions (but branches must have matching types)
* Row Filters as a logical column
* Number, Text, and Logical constants

## Merging/Joining tables

* + Power Query will generate a nested join ([Table.NestedJoin](https://docs.microsoft.com/en-us/powerquery-m/table-nestedjoin); users can also manually write [Table.AddJoinColumn](https://docs.microsoft.com/en-us/powerquery-m/table-addjoincolumn)). Users must then expand the nested join column into a non-nested join ([Table.ExpandTableColumn](https://docs.microsoft.com/en-us/powerquery-m/table-expandtablecolumn), not supported in any other context).
* The M function [Table.Join](https://docs.microsoft.com/en-us/powerquery-m/table-join) can be manually written directly to avoid the need for an additional expansion step, but the user must ensure that there are no duplicate column names amongst the joined tables (because ADF does not support them).
* Supported Join Kinds: [Inner](https://docs.microsoft.com/en-us/powerquery-m/joinkind-inner), [LeftOuter](https://docs.microsoft.com/en-us/powerquery-m/joinkind-leftouter), [RightOuter](https://docs.microsoft.com/en-us/powerquery-m/joinkind-rightouter), [FullOuter](https://docs.microsoft.com/en-us/powerquery-m/joinkind-fullouter)
* Both [Value.Equals](https://docs.microsoft.com/en-us/powerquery-m/value-equals) and [Value.NullableEquals](https://docs.microsoft.com/en-us/powerquery-m/value-nullableequals) are supported as key equality comparers

## Group by

(corresponding M function: [Table.Group](https://docs.microsoft.com/en-us/powerquery-m/table-group)):

* Must be used with an aggregation function
* Supported aggregation functions: [Table.RowCount](https://docs.microsoft.com/en-us/powerquery-m/table-rowcount), [List.Sum](https://docs.microsoft.com/en-us/powerquery-m/list-sum), [List.Count](https://docs.microsoft.com/en-us/powerquery-m/list-count), [List.Average](https://docs.microsoft.com/en-us/powerquery-m/list-average), [List.Min](https://docs.microsoft.com/en-us/powerquery-m/list-min), [List.Max](https://docs.microsoft.com/en-us/powerquery-m/list-max), [List.StandardDeviation](https://docs.microsoft.com/en-us/powerquery-m/list-standarddeviation), [List.First](https://docs.microsoft.com/en-us/powerquery-m/list-first), [List.Last](https://docs.microsoft.com/en-us/powerquery-m/list-last))

## Sorting

(corresponding M function: [Table.Sort](https://docs.microsoft.com/en-us/powerquery-m/table-sort))

* Reducing Rows: Keep and Remove Top, Keep Range (corresponding M functions, only supporting counts, not conditions: [Table.FirstN](https://docs.microsoft.com/en-us/powerquery-m/table-firstn), [Table.Skip](https://docs.microsoft.com/en-us/powerquery-m/table-skip), [Table.RemoveFirstN](https://docs.microsoft.com/en-us/powerquery-m/table-removefirstn), [Table.Range](https://docs.microsoft.com/en-us/powerquery-m/table-range), [Table.MinN](https://docs.microsoft.com/en-us/powerquery-m/table-minn), [Table.MaxN](https://docs.microsoft.com/en-us/powerquery-m/table-maxn))

## Notable Unsupported Functionality (not exhaustive; subject to change)

* Merge columns (however the same effect can often be achieved by clever use of AddColumn)
* Split column (also can often be worked around, albeit trickier)
* Append queries
* Changing Column Types
* “Use first row as headers” or “Use headers as first row”