Transformations of Mapping Data Flow

Azure Data Factory #MappingDataFlows Azure Synapse

Activity	Description	SSIS equivalent	SQL Server equivalent
Multiple inputs/outputs			

New branch	Create a new flow branch with the same data	Multicast (+icon)	SELECT INTO
			SELECT OUTPUT

Join	Join data from two streams based on a condition	Merge join	INNER | LEFT | RIGHT JOIN, CROSS | FULL OUTER JOIN

Conditional Split	Route data into different streams based on conditions	Conditional Split	SELECT INTO WHERE condition1
			SELECT INTO WHERE condition2
			CASE ... WHEN

Exists	Check the existence of data in another stream	Lookup / Merge Join	SELECT * FROM Table
			WHERE **EXISTS**(SELECT ...) | JOIN
			| **NOT EXISTS**

Union	Collect data from multiple streams	Union All	SELECT col1a UNION (ALL)
			SELECT col1b

Lookup	Lookup additional data from another stream	Lookup	LEFT | RIGHT JOIN
Schema modifier			
Cast	Cast columns to different types	Derived Column	SELECT **CAST**(NumStr as INT) as Number, **CONVERT**(DATE, '14/07/2022', 103)
Derived Column	Compute new columns based on the existing once	Derived Column	SELECT Column1 * 1.09 as NewColumn

Select	Choose columns to flow to the next stream	OUTPUT in components, mapping columns	**SELECT** Column1, Column4
			FROM Table

Aggregate	Calculate aggregation on the stream	Aggregate	SELECT Year(DateOfBirth) as Year,
			MIN(), MAX(), AVG()
			GROUP BY Year(DateOfBirth)

Surrogate Key	Add a surrogate key column to output stream from a specific value	Script Component	SELECT ROW_NUMBER()
			OVER(ORDER BY n ASC) AS R#, n
			FROM sys.databases
			+ Incremental Primary Key (with limited capabilities)

Pivot	Pivots row values into columns, groups columns and aggregates data	Pivot	SELECT rowCol, c1, c2 FROM (SELECT sourceCols FROM Table) **PIVOT** (SUM(sumCol) FOR col IN (...))

Unpivot	Unpivots columns into row values and ungroups columns	Unpivot	SELECT rowCol, col, X FROM (
			SELECT rowCol, c1, c2 FROM pvt)
			UNPIVOT
			(X FOR col FROM (c1, c2)) AS unpvt

Window	Aggregates data based on a window and joins with original data	[Sort] + Custom Script	SELECT fun() **OVER**(**PARTITION BY** pc ORDER BY oc) newc, pc, oc, otherCols FROM Table
Rank	Generates an ordered ranking based upon sort conditions specified by the user	[Sort] + Custom Script	SELECT **ROW_NUMBER() OVER** (**PARTITION BY** pc ORDER BY oc) newc, pc, oc, otherCols FROM Table
External Call	Enables custom function by calling out to external REST endpoints row-by-row	Script Component	Not available out of the box. Some tricks are possible: OLE Automation, CLR function, although not recommended.

Transformations of Mapping Data Flow

Azure Data Factory | #MappingDataFlows | Azure Synapse

Activity	Description	SSIS equivalent	SQL Server equivalent
Formatters			

Flatten	Takes array values from hierarchical structures such as JSON (denormalization)	Script Component or
		3rd Party Component
		(ZappySys SSIS JSON Parser Transform)
			`SELECT * FROM OPENJSON(json) WITH (`
			` id INT 'strict $.id',`
			` Name NVARCHAR(50) '$.info.name');`

Parse	Parses columns in document form data (JSON, XML, delimited text)	Script Component or
		3rd Party Component
		(ZappySys SSIS JSON Parser Transform)
			`SELECT * FROM OPENJSON ...`
			`SELECT * FROM OPENXML ...`
			`SELECT value FROM`
			`STRING_SPLIT('clothing,road', ',');`

Stringify	Turns complex data types into strings	Script Component	
Row modifier			

Filter	Filter rows in the stream based on a condition	Conditional Split	`SELECT * FROM Table`
			`WHERE [Column] LIKE '%pattern%'`

Sort	Order data in the stream based on column(s)	Sort	`SELECT * FROM Table`
			`ORDER BY [Column] ASC`

Alter Row	Set action policy on rows when database is sink	Conditional Split
		+ *n* Destinations
			`MERGE |`
			`INSERT, UPDATE, DELETE`
			`IF | WHERE`

Assert	Enables you to build custom rules for data quality and data validation	Conditional Split or
		DQS Cleansing
			`WHERE, REPLACE, SUBSTRING,`
			`IF, RAISERROR, THROW`

Source / Destination			
Source	Source for your data flow. Obligatory first element of every Data Flow in ADF	OLE DB Source and more …	`SELECT * FROM SourceTable`
Sink	Destination for your data flow stream	OLE DB Destination and more…	`INSERT INTO TargetTable`

Version: 2.1 (Aug 2022)