**Azure Data Factory Datasets and Pipelines.**

# Datasets

1. **Excel : DS\_Gen2\_Excel\_Generic**

This dataset is for the excel files stored in the input folder which will be processed through pipeline. It contains two parameters for excel file name and excel sheet index which will be passed through the pipeline.Graphical user interface, text, application, Word

Description automatically generatedGraphical user interface, application

Description automatically generated

1. **Azure SQL Database : DS\_AzureSQLDB\_Generic**

This dataset is the connection to Azure SQL Database. It has two parameters for table name and schema name which will be passed through the pipeline.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface

Description automatically generated

1. **DelimitedText : DS\_ArchiveFolderGen2**

This dataset points to archive-folder which holds the files from input folder which will be processed through pipeline. It contains a parameter for excel file name which will be passed through the pipeline.

It will create directories inside the archive folder for current Year, month and date and will organise the processed files accordingly.

Graphical user interface, text, application

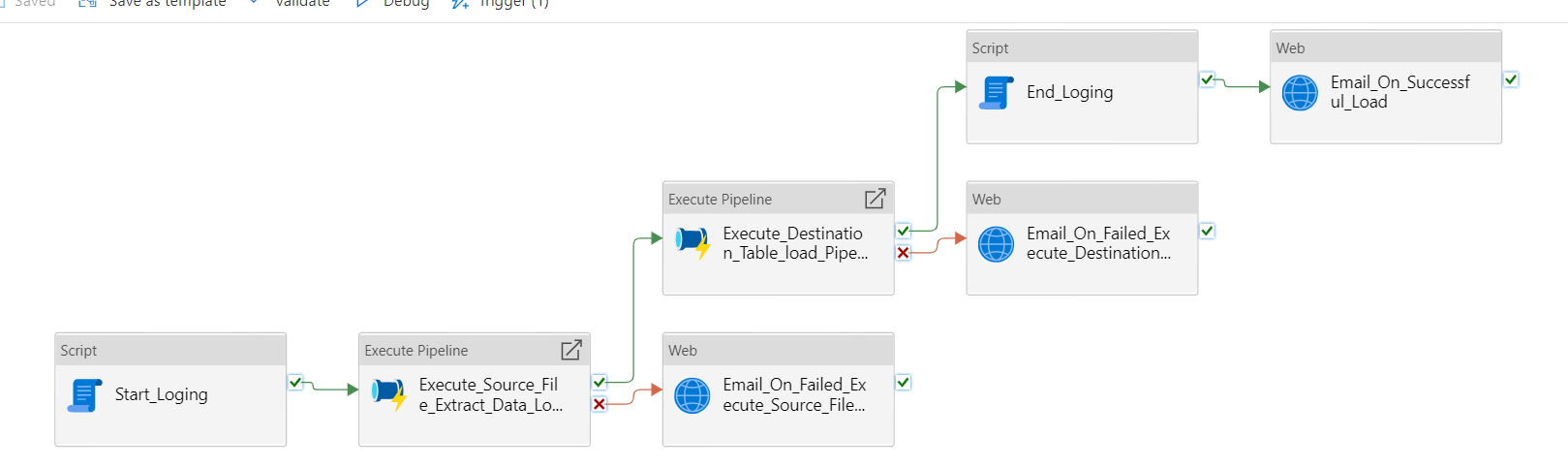
Description automatically generated

Graphical user interface, text, application

Description automatically generated

# Pipeline “ Master\_Pipeline “

This pipeline is the master pipeline will execute both the child pipelines sequentially to complete the entire cycle of loading the data from source files to staging tables and staging tables to Main (DIM and FACT) tables .

If any failure occurs while execution of both the pipelines, an email will be trigerred stating the failure. When both the pipelines are successfully executed, an email will be trigerred stating successful loading of files.

1. **Script : Start\_Loging**

This task will insert the details of current pipeline into etl.t\_ETL\_Log\_Detail table and mark its status as 'Data Load InProgress'

Graphical user interface, text, application

Description automatically generated

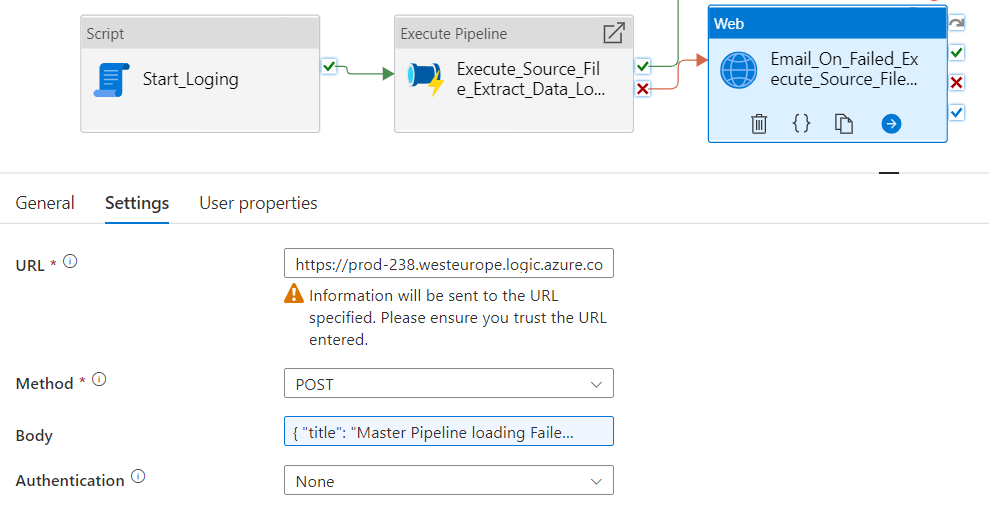
1. **Execute Pipeline: Execute\_Source\_File\_Extract\_Data\_Load\_Pipeline**

This task will execute the child pipeline “[Source\_File\_Extract\_Data\_Load](#_Pipeline_\“_Source_File_Extract_Data)” which is responsible for loading the data from Excel files to staging tables.

Graphical user interface, text, application, email

Description automatically generated

1. **Web: Email\_On\_Failed\_Execute\_Source\_File\_Extract\_Data\_Load**

This web activity will trigger an email if the execution of source pipeline fails. 

1. **Execute Pipeline: Execute\_Destination\_Table\_load\_Pipeline**

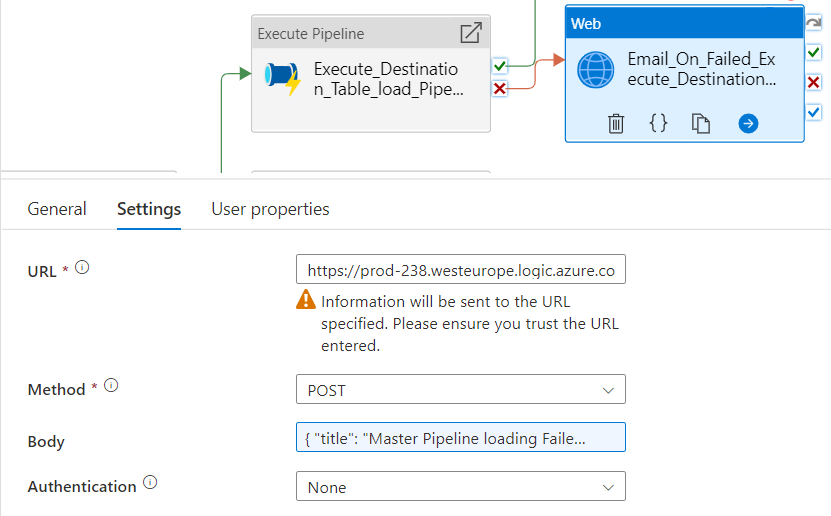
This task will execute the child pipeline “[Destination\_Table\_load](#_Pipeline_\“_Destination_Table_load)” which is responsible for loading the data from staging tables to the DIM and FACT tables.

Graphical user interface, text

Description automatically generated

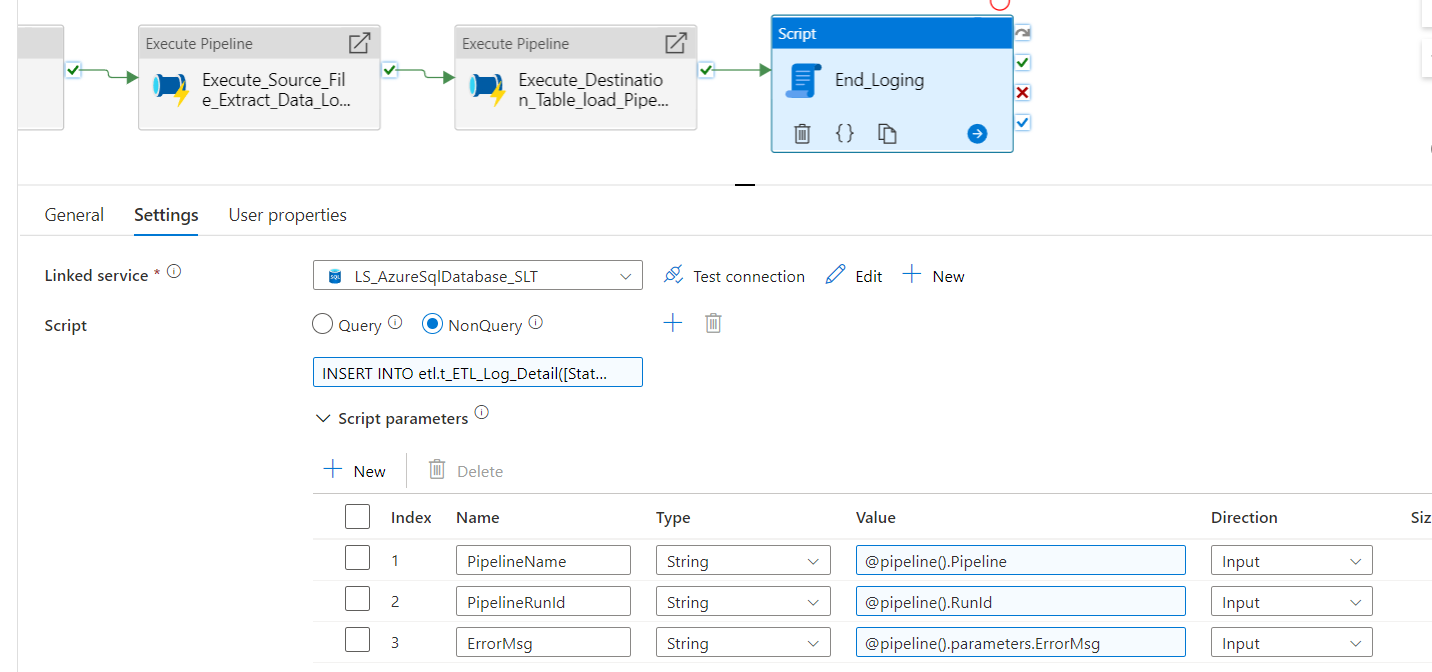
1. **Email\_On\_Failed\_Execute\_Destination\_Table\_load\_Pipeline**

This web activity will trigger an email if the execution of destination pipeline fails.



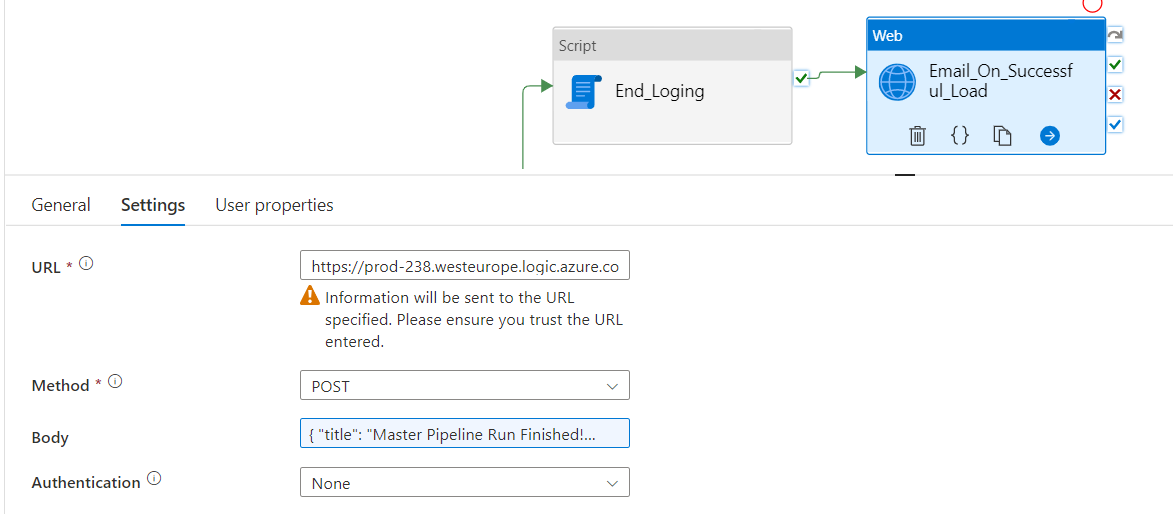
1. **Script : End\_Loging**

This task will update the details of current pipeline into etl.t\_ETL\_Log\_Detail table and mark its status as 'Data Load Completed'.



1. **Web: Email\_On\_Successful\_Load**

This web activity will trigger an email when the execution of both the pipelines is successful and data has been loaded into the final tables.



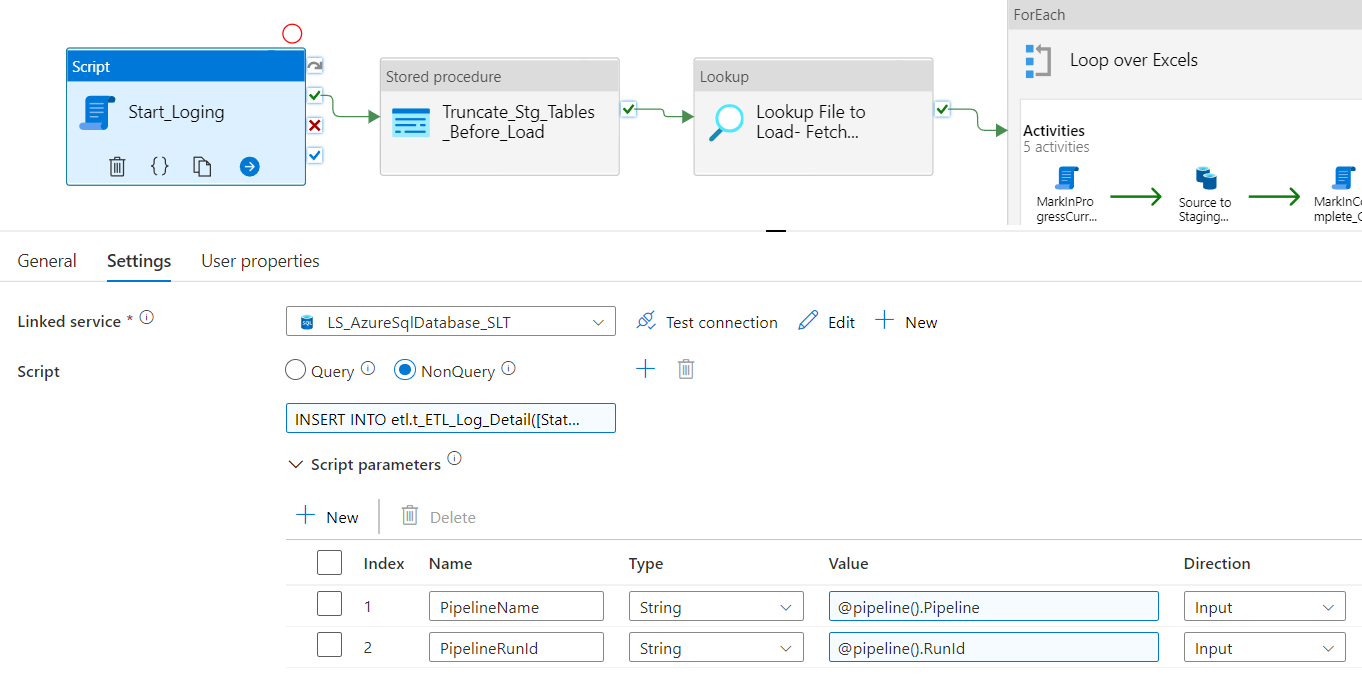
# Pipeline “ Source\_File\_Extract\_Data\_load “

This pipeline loads the data from excel files to staging tables.A screenshot of a computer

Description automatically generated

1. **Script : Start\_Loging**

This task will insert the details of current pipeline into etl.t\_ETL\_Log\_Detail table and mark its status as 'Data Load InProgress'



1. **Stored procedure : Truncate\_Stg\_Tables\_Before\_Load**

This task will execute a SP which is responsible for truncating all the staging tables which need to be loaded again.

Graphical user interface, text, application

Description automatically generated

1. **Lookup : Lookup File to Load- Fetch MetaData**

This task will check using a sql query if the filename in etl.t\_ExcelMetadata which needs to be loaded (where ImportFlag=1 and load date is less than current date).

Graphical user interface, application

Description automatically generated

1. **ForEach : Loop over Excels**

This loop will process all the matched files one by one.

Graphical user interface, text, application, email

Description automatically generated

* 1. **Script : MarkInProgressCurrentFile**

This task will update the table etl.t\_ExcelMetadata with status as 'InProgress'

and LoadStart\_Dt as current timestamp for the current file being loaded.

Graphical user interface, application

Description automatically generated

* 1. **Copy data : Source to Staging Tables**

This copy data task will pick the required excel file and its sheet index and load the data from it to the corresponding staging table.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

* 1. **Script : MarkInComplete\_CurrentFile**

This task will update the table etl.t\_ExcelMetadata with status as 'Completed'

and Load\_Date as current timestamp for the current file loaded.

Graphical user interface, application

Description automatically generated

* 1. **Copy data : Copy File from Input to archive**

This task will copy the processed file from Input\_Folder to Archive\_Folder by concatenating filename with current timestamp.

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

* 1. **Delete : Delete from Input Folder**

This task will delete the processed file from Input\_Folder.

Graphical user interface, application

Description automatically generated

1. **Script : End\_Loging**

This task will update the details of current pipeline into etl.t\_ETL\_Log\_Detail table and mark its status as 'Data Load Completed'.

A screenshot of a computer

Description automatically generated

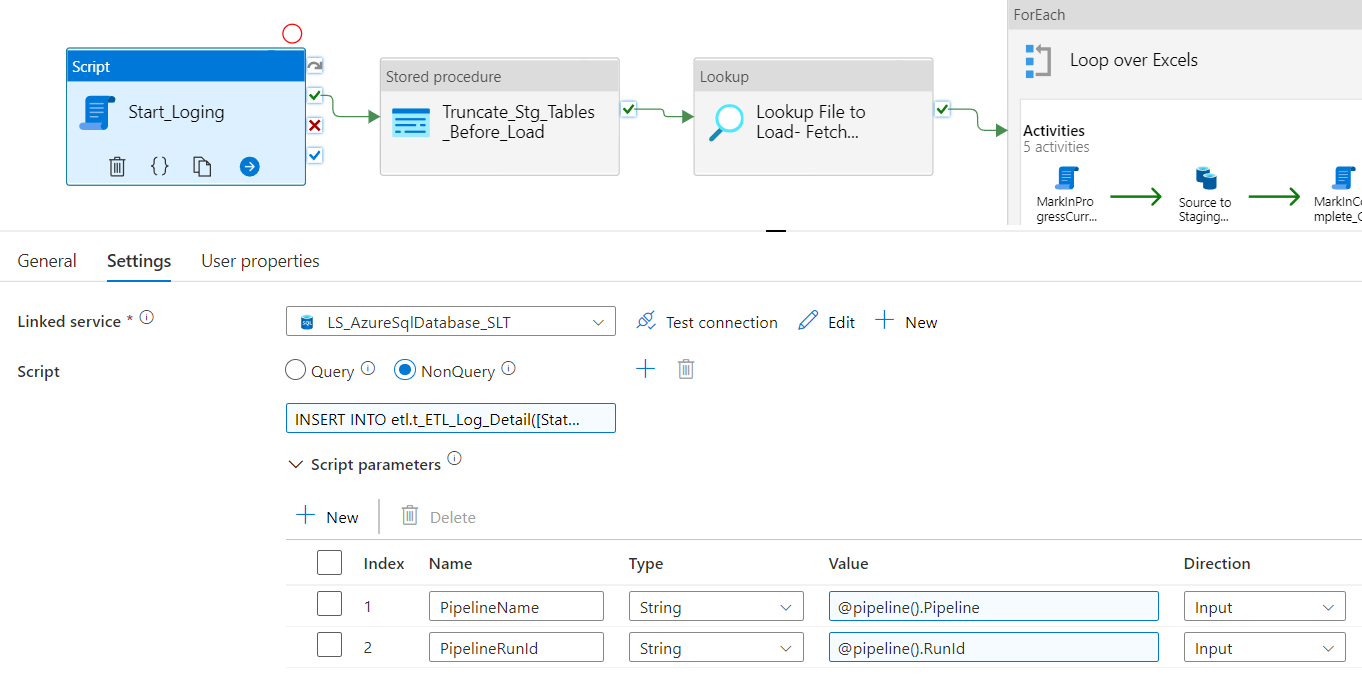
# Pipeline “ Destination\_Table\_load “

This pipeline loads the data from staging tables to main destination tables ( DIM and FACT). Graphical user interface

Description automatically generated

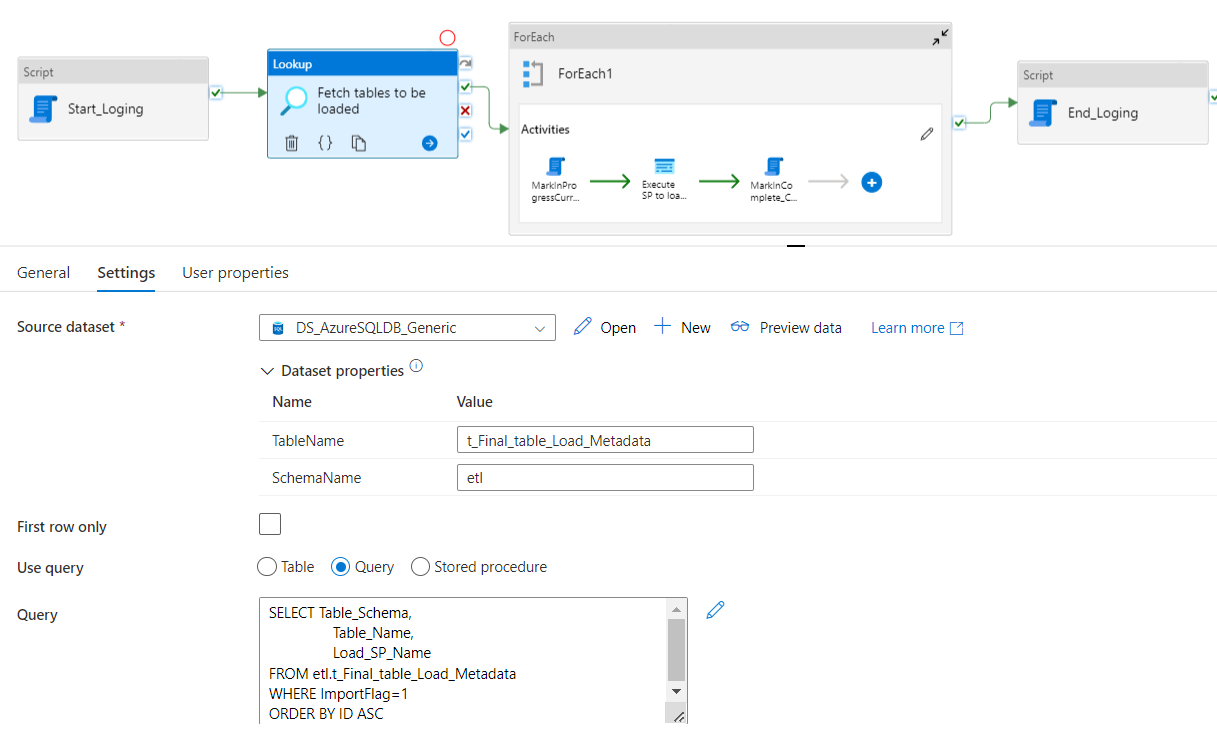
1. **Script : Start\_Loging**

This task will insert the details of current pipeline into etl.t\_ETL\_Log\_Detail table and mark its status as 'Data Load InProgress'



1. **LookUp : Fetch tables to be loaded**

This task will check using a sql query for the details of the tables to be loaded in etl.t\_Final\_table\_Load (where ImportFlag=1).



1. **ForEach : Loop over Excels**

This loop will process all the tables one by one.

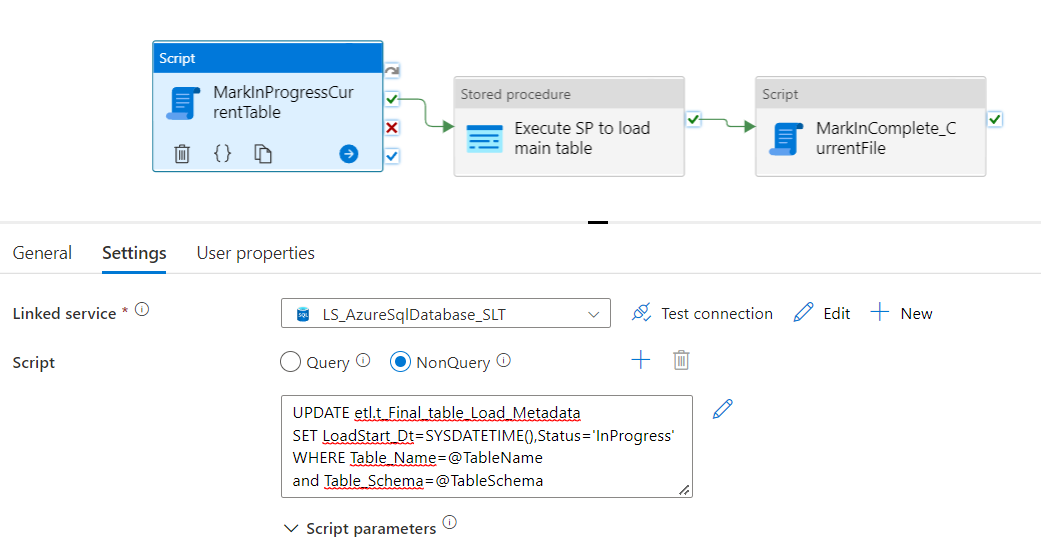
Graphical user interface, text, application, email

Description automatically generated

* 1. **Script : MarkInProgressCurrentFile**

This task will update the table etl.t\_Final\_table\_Load\_Metadata with status as 'InProgress'

and LoadStart\_Dt as current timestamp for the current table being loaded.



* 1. **Stored Procedure: Execute SP to load main table**

This task will pick the required Stored procedure name and load the data using it to the main table.

Graphical user interface, text, application

Description automatically generated

* 1. **Script : MarkInComplete\_CurrentFile**

This task will update the table etl.t\_Final\_table\_Load\_Metadata with status as 'Completed'

and Load\_Date as current timestamp for the current file loaded.

Graphical user interface, text, application

Description automatically generated

1. **Script : End\_Loging**

This task will update the details of current pipeline into etl.t\_ETL\_Log\_Detail table and mark its status as 'Data Load Completed'.

Graphical user interface, application

Description automatically generated