

## Assignment

Rabobank receives monthly deliveries of customer statement records. This information is delivered in CSV format. These records need to be validated.

We would like to solve this by using Azure Synapse.

Infrastructure As Code (If this is feasible, if not please let us know on how we can evaluate the working and outcome on this assignment):

- Storage Account
- Azure SQL database
- Synapse Workspace

## Input

The format of the CSV file is a simplified format of the MT940 format. The format is as follows:

<i>Table 1. Record description</i>	
Field	Description
Transaction reference	A numeric value
Account number	An IBAN
Start Balance	The starting balance in Euros
Mutation	Either an addition (+) or a deduction (-)
Description	Free text
End Balance	The end balance in Euros

## Output

Create a pipeline (preferably using PySpark) in Azure Synapse to retrieve the CSV file from a Storage Account and do the validations.

There are two validations:

- all transaction references should be unique
- the end balance needs to be validated

At the end of the processing, the values are written on the database and a report must be created. The report shows the transaction reference and description of each failed record.

**Optional:** Expose values via an API report.