M8T4 – Data pipeline exercise part 4

Next Phase of Design Based on Scottish Data

Your task is to update the design of your database schema to accommodate the data from the Scottish version of the software application.

The first step will be to update the database schema.

The second will be to re-implement the pipeline to clean, validate and load the four data sets you have received, into your revised database schema.

To help you with this the Scottish application team has provided you with some additional test data. It consists of the data for ten Scottish users plus data about when they logged into the website to access the application.

Your instructor will again act as the stakeholder SME and will answer any questions that you may have about this new data.

This is the third of the additional phases involving other countries supplying you with data from their localised version of the application and you need to accommodate the different formats and requirements in your schema and pipeline.

Remember that you must create full documentation for both the schema and the pipeline so that other teams can implement it.

The 10 sample records from the Scottish subsidiary are in the *SC User Data.csv* file.

The audit file showing the logins for these 10 users in January 2025 is called *SC-User-LoginTS.csv*. It gives the username and the Epoch timestamp of each of their logins during the month.

BPP