

L5DE M8T1 - Data pipeline exercise 1

Briefing - Initial Design Based on UK Data

Your task is to design a database to receive the data about users coming from an existing software application.

The first step is to design the database schema that will receive the data.

The second will be to implement a pipeline to clean, validate and load the data into your new database schema.

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

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

In subsequent phases additional countries will supply you with data from their localised version of the application.

You will need to accommodate their different formats and requirements in your schema and pipeline as the project evolves.

You will need to establish full documentation for both the schema and the pipeline that can be handed to another team so that they can also implement it.

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

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