Scenario
At Schoolzilla, we want to help people use data to change students' lives for the better. But sometimes the data isn't perfect, and needs a lot of cleanup before certain reports can be useful.

For example, imagine the only version of data a customer has for a recent test is in a spreadsheet in the following format.

Student Number	Math Score	Science Score
1597530	100	80
2468975	85	
8675309	blue	95

As you can see, the spreadsheet has missing values in some important columns, a color is in a column that is supposed to have the test score, etc. And it turns out that Student Number 8675309 isn't a real student number.

We also need to reshape this data from it's current format to prepare it for loading into the data warehouse. The table that you would be loading has similar purpose but a different structure:

student_id	subject	score
1597530	Math	100
1597530	Science	80
2468975	Math	85

You've accepted the challenge to build a tool to help our users take data in whatever format they have it in, then perform the required cleanup and mapping on our website. The types of things your tool should be able to do include:

- map columns from the customer dataset to the columns required for the Schoolzilla data model
- display sample rows from the customer dataset
- allow values to be changed by the customer
- validate that the values entered are appropriate for the Schoolzilla element they've been mapped to

Question 1
How would you design a solution to this problem?

(These bullets are just some thoughts that may help you get started, but you should answer the question how you think most appropriate.)

- Define the challenges you think need to be overcome.
- Describe the main components of your design and how they come together to solve the problem.
- You may want to note any tools, libraries, patterns, etc, you think are essential to incorporate into your design and explain why.
- You can communicate your design in whatever format(s) you think appropriate (psuedocode, real code (in your preferred language), diagrams, etc).
- You do **NOT** need to implement a working solution. And can make some simplifying assumptions in order to complete the task within a few hours as long as your assumptions are reasonable and clearly stated.
- As a guideline for level of detail, imagine you are presenting your thoughts to technical teammates in an early stage design review. Provide enough detail to show you've thought through the challenges, but no expectation that it's a fully baked solution.

Question 2
How would you verify that your solution continues to perform well?

(The bullets are just some thoughts that may help you get started, but you should answer the question how you think most appropriate.)

- Describe some of the most essential unit tests you'd create to ensure that your code keeps working smoothly as development progresses?
- How would you determine which components are performance bottlenecks?
- What if anything changes if you suddenly had to handle 20x the traffic with the same performance, or files that were extremely large datasets?

If you have any questions about the homework, please don't hesitate to ask.