Round 2

About
This assignment is a part of screening process for candidates who applied for a role at our organization.

Data
You are being provided with the required source data in file 'customers.csv' & 'orders.csv'. You can take twitter data through Twitter API.

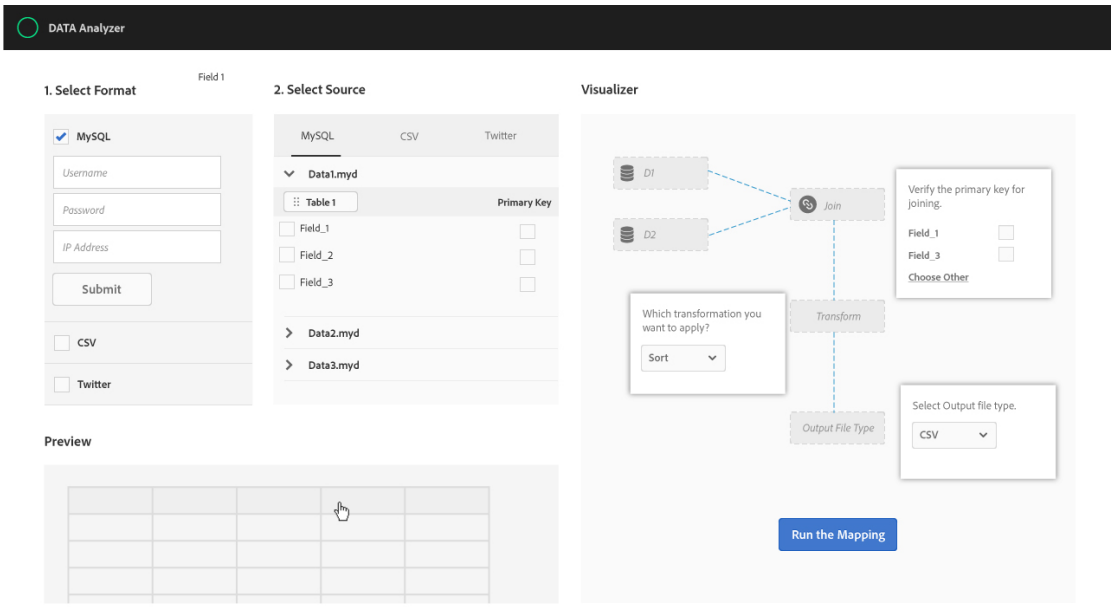
Background
The assignment focuses on development (design is provided, hence not required) data pipeline on a web application that can dynamically extract, transform, & load data from source to target data layer.

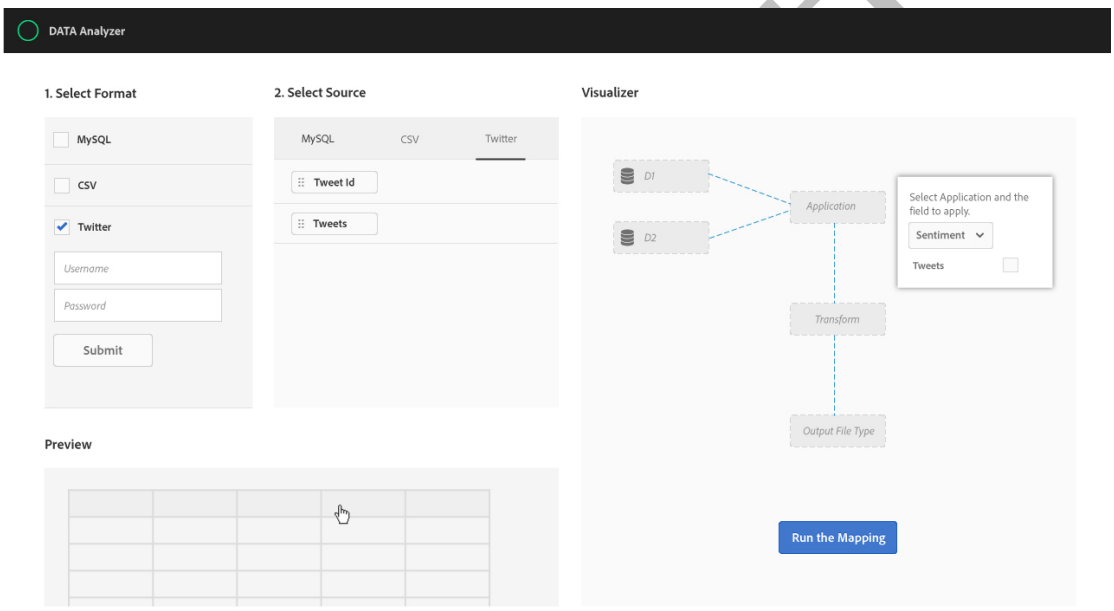
Problem
You need to implement from scratch in development stack of your choice as per design description below.

Backend
Create connectors for MYSQL and CSV to get data from any instance. Get the data (UserID & respective Tweets for keyword 'demonetisation') for Twitter through its provided APIs.
Second step is to select the fields where the transformation has to be applied.

Frontend
Shown above are the screens of a 'Data Analyzer' Application, which acquires data from MySQL, CSV formats, & twitter data with the help of twitter API.
First step is to select data format(s) from which the data has to be imported after entering proper credentials for MySQL, CSV, & Twitter and for simple 'browse' option can be used for importing tables.

Refer wireframe files Data_Analyzer_1.jpg & Data_Analyzer_2.jpg
(pasted below)





For MySQL, Data →Table→Fields flow is shown in the window where fields and primary key (CustomerID) is selected. Take customers.csv and import into your MYSQL database as CUSTOMERS table.

After selecting, table (having fields one or many selected) is dropped in the 'Visualizer' window and similarly from above procedure this table can be dragged & dropped. (Current screen shows placeholder for two drop points but it can be one or many).

After dropping the tables in the 'visualizer' window, now drag & drop them to the 'join' placeholder. Now verify the primary key for joining.

Next drag & drop the joined table to the' transform' window, where the system prompts to apply transformation to the fields (Sort by – alphabets). Then choose the format for the output (MySQL or CSV).

'Run the mapping' button opens a new window showing the full table having all the transformations which was designed in the mapping.

For CSV format, after browsing the CSV file(s), repeat the similar procedure (as stated above) of selecting the fields and primary keys to apply transformations. Take orders.csv as ORDERS table.

Fields from CSV and MySQL can also be joined in single mapping.

For Twitter data, only table with two fields is made and the fields are 'UserID' and 'Tweet'. Drag them to the visualizer and then to the 'Application' placeholder where the system asks the algorithm to apply on the 'tweets' field, like 'unique word count' which provides count of unique words. After applying the algorithm for 'unique word count', apply filter transformation to get tweets which have more than five unique word counts.

After every transformation/joining/etc taking place in the Visualizer, the resulting table (only first 20 rows) is previewed in real-time in the 'Preview' window.

Your Task:

1. Create connectors for MySql, CSV and Twitter.
2. Create pipelines to Extract, Load, and Transform the data as per the instructions.
3. Create frontend as closely as possible to the application as discussed.

Submission

You are required to submit application in a zipped file which can be hosted on WAMP.

Evaluation

You shall be scored on the following:

1. Look & Feel
2. Functionality
3. Performance