Hello,

For the past week, I have been working on this project wherein I am taking in the unstructured data (.json files) and creating a new structured relational model from it. Following that, I queried this newly created relational schema to answer certain business questions.

(For your reference, I have also attached the new relational schema).

But while working on this project, I had a few questions and came across certain issues for which I want your feedback.

Questions
 1. What's the source of this unstructured data? How frequently the data was fetched from these sources?
2. I observed a lot of missing data. So it would be interesting to see that this is because the data was never recorded or there's some issue when the unstructured data is being migrated into the data source?

Issues
1. In the table named "user", I found multiple duplicate entries for each user. This makes the results of my certain queries inconsistent.
Hence, in my opinion, the table should contain only one instance per user.
2. In the "brands" table, in my opinion, Brand Code should also be unique for each brand.
3. There are many missing values in different tables which makes the results of certain queries inconsistent.

Additional details regarding the following will also help me in understanding the above-mentioned and certain other issues well.

1. Data Source- More information about where and how frequently the data is fetched.

Speaking of the python script to convert unstructured data into structured data, it's not production-ready yet.

To make it production-ready, I intend to follow the below steps.
1) Creating specific python virtual environments for each separate deployment
2) Installing static code analysis libraries in the virtual environment as dev dependencies
3) Turn any hardcoded values like database credentials into parameters. 4) Building a Docker file to containerize the code.

That's all from my end. Please let me know if you have suggestions to make this process more efficient. Your time and effort are much appreciated.

Best ,
Anuj Raju Chanchlani