* What questions do you have about the data?
* How did you discover the data quality issues?
* What do you need to know to resolve the data quality issues?
* What other information would you need to help you optimize the data assets you're trying to create?
* What performance and scaling concerns do you anticipate in production and how do you plan to address them?

If we are doing the production level ETL or DW then we will need a bit of time in flattening the files because they are not simple. So, my greatest challenge would be flattening the json files. In regard to optimization as I mentioned the receipt json file has two important dictionaries that I would think of them as separate tables so to avoid duplicating receipts for every purchase with more than one brand rewarded.

Also, if we plan to create the tables then I would prefer the barcode in the rewards table to be filled in the background so to avoid missing out. In future once the datasets grow use Common Table Expressions to efficiently handle hierarchical data and for advanced level use cache if the outcomes are being posted to applications.

But as of now since the tables are just four I think there would not be many delays in querying no matter how big the data could be. I think the time for search would grow but no better way to improve in terms of speed