**Q&A**

1. **Lecture 8 shows the deadline is week 12 while canvas shows 31,Oct. Which one should we confirm to?**

According to school calendar from the following link:

https://www.nus.edu.sg/registrar/docs/info/calendar/ay2022-2023.pdf

Oct 31 is the first day of week 12.

You will be notified if the deadline is to be extended.

1. **For data generation part, we have to specify assumptions. I’m not quite sure about this. For example, we can based on 10% of the reviews or try to choose some reasonable reviews. Does it mean we only have to use part of data rather than the whole reviews data?**

You can use 10% review to generate transaction data, for example, if you generate one transaction record based one review, the total number of transaction records would be 10% of total reviews. But for other tasks, you need to use the whole dataset. 10% here is something you need to decide and state in your assumption or specification for data generation.

1. **If reviews cannot provide enough information for us, should we just make up the data randomly by ourselves?**

You need to make up some data for transaction data because reviews won't provide all the information. In this case, it is important for you to specify your assumptions when generating data. For example, the total amount of spending is generated based on a normal distribution with mean and variance as x & y.

1. **I would like to confirm that we only need to use one dataset for the project, and we should be storing the data in both relational and non-relational databases just like we did for the university data in class, not one dataset for the relational database and another one for the non-relational database.**

Use **ONE** dataset for two approaches.

1. **We have a question on DSA5104 Group Project deliverables. From Canvas page, It is mentioned that we need to submit a report that consolidates all the tasks and details required. However, from the lecture, we did hear from you that if the details are encompassed in the PPT slides for presentation in Week 12, then we don't really need to submit on project report as PPT slides alone will suffice. Can we double check with you whether this understanding of ours is correct?**

Project report and slides are different ways of presenting your work, in which presentations of 10-15mins (including Q&A) focus more on the main ideas of your approach, highlights, and important results, while a lot of details and comprehensive descriptions are left in the report.

You may submit slides as your group report if you think it proper and matches **ALL** the requirements.

1. **We are wondering that after BCNF decomposition of the original database into some tables, does it mean that we should reload the files and rebuild database containing tables satisfying BCNF, or we can just use the original database to conduct SQL queries?**

You need to use the tables satisfying BCNF in the relational Databases when conduct queries. But you don't need to maintain BCNF tables in the non-relational database. You can use nested objects (flexible schema) in MongoDB to demonstrate the differences (pros and cons) between relational and non-relational databases.

1. **And there is another question that we are not sure of, which is about the assumed transactions. Do we need to include a transaction table when drawing our ER diagram and relational schema?**

You need to include transaction data when you design your relational database.

1. **And if we need to store the transaction table, do we need to store it in MongoDB?**

Yes, you are suppose to manage the same data in both database systems.