Data Management Plan - Draft Timeline

The Project of comScore, Inc. with K. Company

0-2 Months

- **Database storage optimization**
  - Delete/clean redundant columns
- Testing for comScore
  - Storage saved
- Feedback from comScore
  - Whether the deleted columns would impact on the existing process
  - Whether the efficiency has improved when retrieving the data

3-6 Months

- **Database performance optimization**
  - Set up index for frequently used column(s) to improve efficiency when searching for that column(s)
- Testing for comScore
  - Index storage analyses
    - Extra storage
    - Efficiency improved
- Feedback from comScore
  - Whether too much storage was required by setting up index
  - Whether the optimization on database performance has improved efficiency when retrieving frequently used columns by setting up index

7-9 Months

- **Existing process performance optimization**
  - Stored procedure
    - Revise syntax in queries to improve efficiency
      - i.e. Replace sub-queries with JOIN
      - i.e. SELECT specific column(s) rather than SELECT *
    - Avoid unnecessary temp tables - several JOINs can be done in one temp table instead of using multiple temp tables
- Testing for comScore
  - Stored procedure analyses
    - Storage saved
    - Efficiency improved
- Feedback from comScore
  - Whether the optimized queries and reduced tables have improved the entire efficiency
  - Whether there is any impact on the existing process

10-12 Months

- **Monitoring process build up**
    - Automatically detect data processing error
    - Log the error(s) into the "error table" of the database
    - Set up scheduled email process for error reporting
- Feedback from comScore
    - Whether the monitoring process build up would help for improving the efficiency of finding and fixing errors

Overview

Based on the existing problem of the database, I came up with the plan with the solutions respectively. Due to the huge database that automatically updates with high frequency, the most important thing came to my mind is to improve the overall efficiency of the database in both retrieving and storing. Therefore, I have included the solutions above regarding data storage optimization, data performance optimization, and existing performance optimization. Also, since my client mentioned during the last interview that they have a problem of finding and fixing data processing error, I have also included the part of monitoring process build up in the timeline. On each task, I have included testing on my side (when applicable) and feedback from comScore in order to follow up on the results of the timeline implementation.

As for the data storage optimization, the client mentioned that there might be some redundant columns that haven't been removed, so I think it is necessary to remove all the redundant columns from the database to reduce the storage space needed and get prepared for the task of data performance optimization.

During the data performance optimization, due to the huge and complex database, I would suggest the client to set up index for frequently used columns to help retrieve the data faster when searching for those columns. However, on the other side, I am still not sure whether this would drag down the overall efficiency again as setting up index would potentially require more storage space.

After the index setup, queries and temp tables can be optimized through the process of existing performance optimization. In this part, queries in the automated process of data processing can be simplified by modifying syntax to save the storage and improve the efficiency of data processing; temp tables can be combined to save storage and improve the efficiency when multiple temp tables are not necessary.

Now, at this stage, an automated error detecting function and scheduled emailing function can be done by stored procedure. This monitoring process could help with detecting and fixing data processing error timely.