

CSE 202, Winter Session 2021

Take Home Mid-Sem Exam

Base Marks: 100
Marks: 79

Bonus

Exam Release Date and Time: Friday, March 5@10:00AM or before

Submission Date and time: Monday, March 8@9:00AM.

Instructions:

- This is a group exam. Your group number and its members have been notified to you. One submission per group is required, and therefore, all students in the group will get the equal marks. Submission will be accepted ONLY through Google Form. In case your group members submit the answer sheet multiple times, the first submission will be accepted as the **FINAL** submission.
- It is an OPEN book/material exam. However, we trust you will not take cross-group help and check your answers. **List your assumptions, if any. We will accept all reasonable assumptions.**
- For late submission, 4 marks will be deducted for each 30mts late submission.
- **Good News ☺:** Each group will get 5 bonus marks in this exam if the group submits the answer sheet as a single pdf file through Google form and includes a 'Table of Contents/Answers' [Answer#, Page#] in the first page of your answer sheet. For every sub-part of the question, please make a separate row in the table. The 'Table of Contents/Answers' should look like --

Answer#	Page#
3	6-7
4A	8
...	...

- **'We can' Good News:** Questions 5F and 7C are optional and carry 14 bonus marks in this exam. So, give it a try.
- **Wow ☺☺:** If your group implements the database design as described in Question 1, populate few tuples in tables, and run all SQL statements and embedded SQL program (Questions 5 and 6), you may get up to **60 bonus marks in this exam**, subject to condition that you demonstrate the running system by **March 12** to any of your group member TA.
 - Creation of Tables with all constraints and data population successfully – 25%
 - Execute SQL statements successfully – 15%
 - Execute embedded SQL program successfully – 10%
 - Define Views and Grant permission to all stakeholders (Participants, Members, Group Leaders, Distributors, Outsiders) for accessing information relevant to them – 10%
- If you want to join this one-time 'Wow' club, send an email to any of your group member **TA by March 10** and block his/her calendar for evaluation. Again, the evaluation needs to be completed by 11:59PM of March 12. Thus, the onus is on you to schedule the meeting. If any problem, get in touch with TF.
- The weightage of the bonus marks in this exam will be the same as the weightage of the base marks. That is, the overall weightage of this exam is 15%.
- Ideally, you have an opportunity to get up to 179 marks out of 100. Good luck to you.

Question 1: Let us design a DB for the following requirements:

- A highly popular 'McM Sangeet company' plans to organize a competition for searching the right talents to train them for creating music albums. The company advertises their recruitment requirements on different channels (both print and digital media) for inviting talents by submitting their personal information, prior experience, and a 2-5 minutes media file, which is of either an audio file (for songs) or a video (for songs and/or playing music instruments), to a given URL in the advertisement. The company writes in the advertisement that the shortlisted candidate will be informed by Phone and/or Email.
- A candidate can submit more than one entry and has an option to provide more than one phone number in the submission, which are unique to him/her.
- After the closing date of entry submission, a panel evaluates all entries and recommends a set of candidates to invite them for the next round. The company maintains the information about the panellist (personal information, industry experience, association with the McM company). A candidate can see the outcome of their application online.
- These shortlisted candidates are then invited to perform a 'Live' show in Mumbai and finally top ' n ' candidates are selected in each album category (i.e. audio and video). These $2n$ candidates are called as the member of McM-2020.
- Different music groups (pop, classic, leisure, evergreen, ...) are formed to create the music albums (audio/video). Every member belongs to one or more music group which is moderated by a director who himself/herself is a member of the group. Each member has a different role to play in each album.
- Once the album is created and approved by McM Director, its trailer is released online for limited time to the outsiders to give their comments (like and dislike), and all this recorded in the database (album name, date of release, number of visits, number of likes/dislikes, ...).
- The McM decides the price of the album after analysing the data collected from its trailer release, and then the album is released to distributors who eventually will sell it online. Note, each distributor may be charged a different price per unit depending on the negotiation between the McM and distributor.
- When a distributor sells an album, the download request comes to the McM site. Thus, the McM company maintains the record for each download (Incoming URL – identifying the distributor, Album#, Date, Download Status – success/failure) so that they can track the number of downloads for raising the invoice.

Design and draw an ER diagram. List your assumptions, if any. We will accept all reasonable assumptions. Your E-R diagram should clearly --

- Identify all entities (Sole/Generalized/Weak...) and their attributes (atomic, multi-valued, derived) (6)
- Identify relationships between these entities and their attributes (6)
- Identify and underline the primary key of each entity (4)
- Identify entity relationship participation and constraints (4)

Question 2: Convert your E-R diagram of Question 1 into relational schemas. Underline the Primary Key and list all candidate keys in each relational schema. (8+2+2)

Question 3: Has your group end up making a 'good design' of the McM Sangeet database in Question 2? If not, why? If yes, why? (1+5)

Question 4: Write the Relational Algebraic Expression (RAE) for each of the following query on relational schemas of Question 2.

- A. List all 'Audio' albums released in 2020. (2)
- B. List all participants who have submitted both Audio and Video files. (2)
- C. List all members who have been the member of more than one group (3)
- D. List all members of 'Pop' music group who are not part of any other music group. (3)
- E. List all distributors who sold all types of albums (4)

Question 5: Write the SQL statement on relational schemas of Question 2 for

- A. List all 'Audio' albums released in 2020. (2)
- B. List all members who have been the member of more than one group. (4)
- C. List all members of 'Pop' music group who are not part of any other music group. (4)
- D. List all participants who have submitted both Audio and Video files. (4)
- E. The McM company would like to analyze the data for "Which advertisement channel has been effective that attracted maximum number of entry submissions?" (4)

Bonus marks

- F. The McM company would like to recognize the distributors who have sold maximum number of albums in each category (Audio/Video) in the year of 2020, and also would like to award all members involved in creating these albums. Write an embedded SQL program to generate the report listing these member/distributor names with albums. (10)

Question 6: In the above DB, the McM Sangeet Company would like to store the aggregate information about '**total entries/submissions**', by (1) State (i.e. Delhi, UP, MP, WB, ...) of the candidates (2) Type of Albums submitted (Audio, Video) (3) Age group of candidates (15-20, 21-25, 26-30, 31-40, 40-100).

- A. How will you model this data? Describe it crisply and precisely. (3)
- B. Write the SQL statement to store and analyze the data by state, type of albums, age group. (3)
- C. The company would like to analyze the data by zone (North, South, East, South), where each state belongs to only one zone. Write the SQL statement for the same. (3)

Question 7: A student ends up making the following relational schemas for the McM Sangeet DB. Definitely this is not a good design.

Participant(Member#, Name, Age, City, Phone, Email, Prior_Experience, Advt_Seen, Album_Type, Submission Date, File_Upload_Path, Status_Round1, Status_Round2)

Panelist_Album_Evaluation(Panelist#, Panelist_Name, Experience, Association_Month, Association_Year, File_Upload_Path)

Member_Group_Album_Trailer(Member#, Member_Name, Group#, Group_Name, Member-Role, Group_Music_Class, Album#, Album_Name, Album_Type, Date_of_Creation, Album_Description, Group_Leader#, Group_Leader_Name, Album_Approver#, Approval_Date, Trailer_Release_Date, Trailer_Release_URL, Incoming_URL_for_View, View_Date, Comments)

Album_Distribution_and_Download(Album#, Album_Release_Date, Distributor#, Distributor_Name, Distributor_Location, Price, Download#, Incoming_URL_for_Download, Download_Request_Date, Downloaded_Album#, Download_Status)

Some of the following FDs are:

Phone → Member#
File_Upload_Path → Member#, Album_Type
Member# → Member_Name, Age, City, Email

Panelist# → Panelist_Name, Experience, Association_Month, Association_Year
Group#, Trailer_Release_Date → Album#
Album# → Group#
Group# → Group_Name, Group_Leader#, Group_Leader_Name

Album# → Album_Name, Album_Type, Date_of_Creation, Album_Description,
Album# → Group_Leader#, Album_Approver#, Approval_Date
Album# → Trailer_Release_Date, Trailer_Release_URL, Album_Release_Date

Distributor# → Distributor_Name, Distributor_Location
Album#, Distributor# → Album_Release_Date, Price

- A. Why is it a 'bad' design? Describe with an example for each anomaly in each relation. (2+3)
- B. Starting from this 'bad' design, how will you end-up to 'good' design? That is, normalize these relational schemas in the order of higher normal forms starting from 1NF. Describe why do you need to go for the higher normal form, if need to be? Show each step during the normalization and list all the keys in the normalized relation. (4*4)

Bonus marks:

- C. Why is *Download#* in **Album_Distribution_and_Download** relation a *surrogate key*? (4)

Note:

- The primary keys are underlined in each relation.
- A participant can provide more than one phone number.
- Status_Round1 and Status_Round2 values are {Pass, Fail}.
- A member may belong to more than one group and may play more than one role in each album.
- Each group can release maximum 1 Album on a given date.
- Group_Music_Class is (Pop, Classic, Leisure, Evergreen, ...) and Album_Type is (Audio, Video).
- The value of attribute *Advt_Seen* is atomic, which is print, digital/TV/Social/...