PES UNIVERSITY	**PES University, Bengaluru** (Established under Karnataka Act No. 16 of 2013)	**UE20CS903**

JULY2021: END SEMESTER ASSESSMENT (ESA)
M TECH DATA SCIENCE AND MACHINE LEARNING_ SEMESTER I

UE20CS903 – DataBases & SQL

Time: 3 Hrs	Answer All Questions	Max Marks: 80

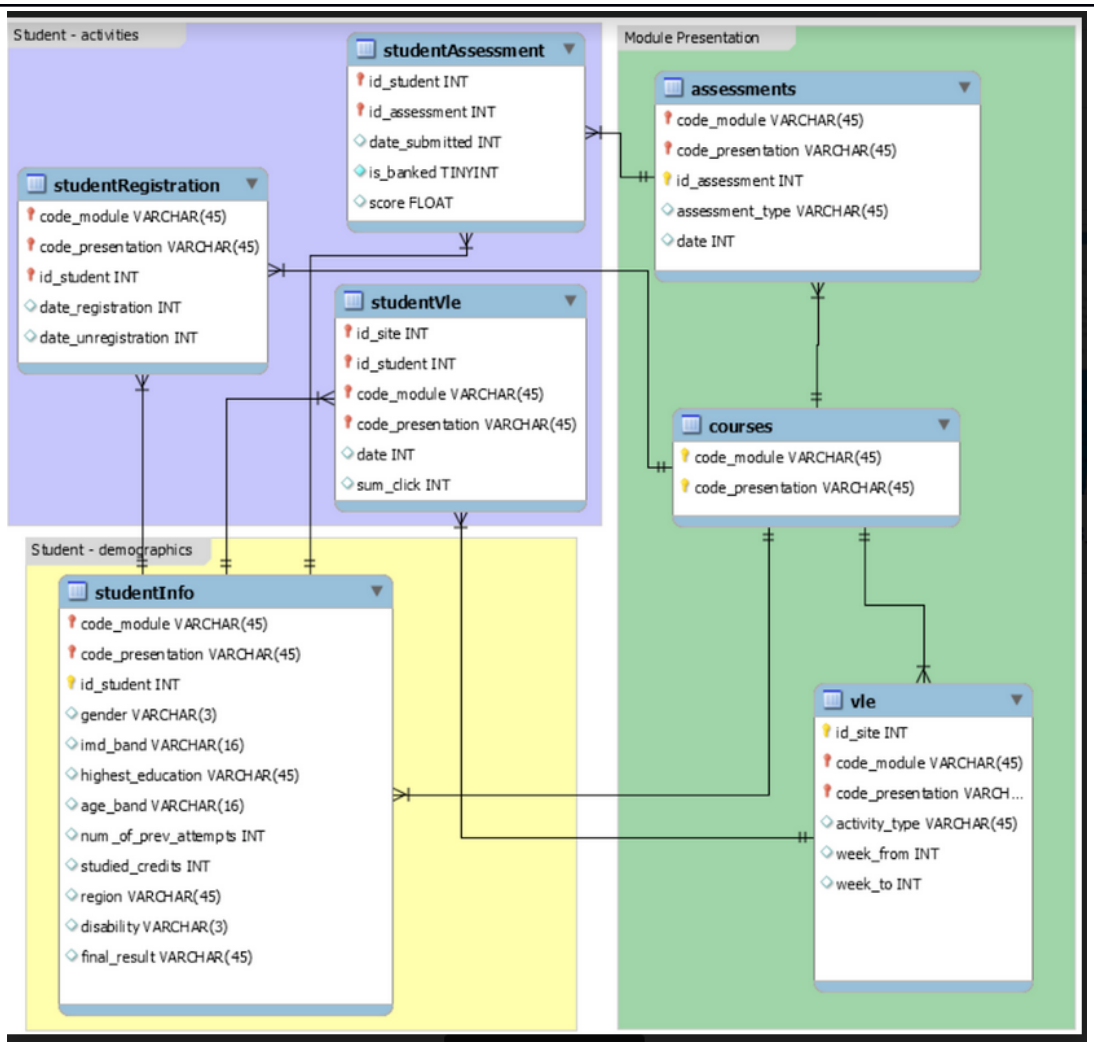
INSTRUCTIONS

- All questions are compulsory.
- Section B and C are coding questions which have to be answered in the system and uploaded in Olympus Login.
- Section A should be handwritten in the answer script provided and signed at the end of the same.

SECTION A – 20 MARKS

1	a	Explain Unique Key with example and describe when to declare an attribute as UNIQUE key instead of PRIMARY KEY.	2
1	b	Write a SQL query to find all duplicate emails in a table named Person. **Note**: All emails are in lowercase. Person Table: +----+---------+ | Id | Email | +----+---------+ | 1 | a@b.com | | 2 | c@d.com | | 3 | a@b.com | +----+---------+	2
1	c	How to add columns to a table in MySQL.Explain with proper example	2
1	d	Given a scenario, an attribute named First Name consists of duplicate values and as well as NULL values. Is it possible to apply Group By clause on First Name?	2
1	e	What is the difference between RANK() and DENSE_RANK()? Justify the answer with sample code.	2
2	a	How do you control the data in the child table when there is a change in the parent table? Explain with an example.	2
2	b	What is a foreign key? How to implement the same in MySQL?	2

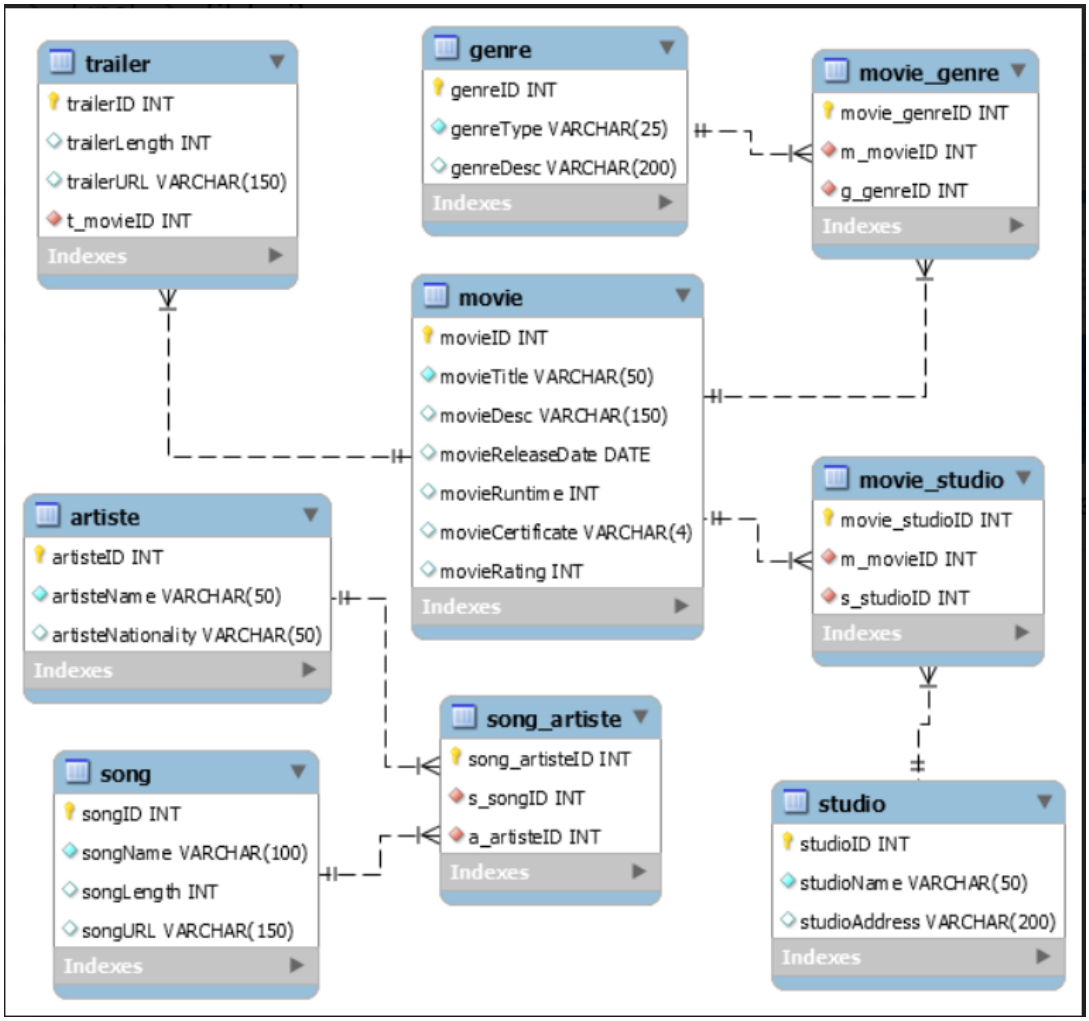
2	c	Write a SQL query to get the second highest salary from the Employee table. +----+--------+ | Id | Salary | +----+--------+ | 1 | 100 | | 2 | 200 | | 3 | 300 | +----+--------+ For example, given the above Employee table, the query should return 200 as the second highest salary. If there is no second highest salary, then the query should return null. +---------------------+ | SecondHighestSalary | +---------------------+ | 200 | +---------------------+	2
2	d	If we drop a table, does it also drop related objects like constraints, indexes, columns, defaults, Views and Stored Procedures? State different types of constraints.	2
2	e	Explain multi-row operators for subqueries with an example.	2
		SECTION B – 30 MARKS	
3	a	**Consider the Open University ER Diagram given below:**	5



Scenario:
Open University is a public British University that also has the highest number of undergraduate students in the UK. It is the largest academic institution in the United Kingdom (and one of the largest in Europe) with 2 Million enrolled students since it was established in 1969. As can be understood from its name, Open University is mainly populated by off-campus students.
This dataset belongs to Open University Online Learning Platform (Also called as "Virtual Learning Environment(VLE)") that off-campus students use for accessing the course content, forum discussions, sending assessments and checking out assignment marks etc. It consists of 7 selected courses (mentioned as modules in the dataset). Different presentations indicated letters "B" and "J" after year for semester 2 and semester 1 respectively.
Additionally, the dataset includes student demographics such as location, age group, disability, education level, gender etc.
Student assessment marks, interactions with the Virtual Learning Environment (VLE) are also included.

		Mission: Personalisation of online education has a big potential to enhance the quality and efficiency of education. Any insights about how people differ from each other in the way we learn would be impressive. Given below are some scenarios, for which you are supposed to use the data and analyse and provide inferences. The inferences can be written as comments in the Query window. A study says that, Students studying more credits at a time(>150) are more likely to withdraw courses than students taking fewer credits. Use the tables to fetch the data and analyse. Write your inferences about the statement using comments. -- (Use Tables:assessments, courses, studentassessment, studentinfo, studentregistration, studentvle, vle)(5marks) -- Sample Output Credits student_withdraw_count More_credits 23 Fewer_Credits 1	
3	b	The University has virtual learners from across locations. The marketing team performing analysis wants to identify the geographic locations with the least number of enrollments. Write a Query to fetch the necessary details for analysis. -- Display region only -- (Use Tables:assessments, courses, studentassessment, studentinfo, studentregistration, studentvle, vle) (8marks)	8
3	c	The University's Academic team wants a report of the list of students who have taken assessments along with the details of modules and assessment_types. In Addition include the list of students who are yet to take up any assessments. Write a query to generate the report. -- (Use Tables:assessments, studentassessment, studentinfo)(5 marks) -- Sample Output: id_student code_module code_presentation assessment_Type date 28061 AAA 2014J TMA 54	5
3	d	Inpatient Charges Dataset consists of hospital charges in the various hospitals in the US for the top 100 diagnoses. The dataset is owned by the US government. This dataset will show you how the price for the same diagnosis and the same treatment and in the same city can vary differently across different providers. It might help you or your loved one find a better hospital for your treatment. You can also analyze to detect fraud among providers.	5

		To understand the number of discharges from every healthcare provider so far, write a query to generate a report such that the provider with least discharges is displayed on top. (Use Tables: InpatientCharges)(5 marks) Sample Output provider_id Provider_Name total_discharges 370089 TAHLEQUAH CITY HOSPITAL 11 210061 ATLANTIC GENERAL HOSPITAL 11	
3	e	To understand the insurance covered ratio for all the hospitalizations, generate a report with the list of providers, Insurance covered ratio in the order of highest coverage to lowest. Insurance coverage ratio is calculated as the average covered costs divided by the total cost which involves average covered charges, total payments and medicare payments. (Use Tables: InpatientCharges)(7 marks) -- Sample Output provider_id Provider_name Insurance_coverage_ratio 670024 NORTH CYPRESS MEDICAL CENTER 87.05475259950786 530010 LANDER REGIONAL HOSPITAL 57.482408438225825	7
		SECTION C – 30 MARKS	
4	a (i)	Consider a table named ‘home’ storing Arsenal football club's performance in the league at home location in 2003-04 season, while the table ‘away’ stores Arsenal's away(other than home location) performance in the same season. Note: that a team is awarded three points for a win, one for a draw and zero for a loss. Win: Number of Goals scored > Number of Goals conceded Draw: Number of Goals scored = Number of Goals conceded Loss: Number of Goals scored < Number of Goals conceded Write a query to determine the number of teams against whom Arsenal has won altogether both in home and away grounds. (Use Table: - home, away)(4 marks)	4
4	a(ii)	Scenario:Consider two tables, storing the date of commencement of tenures of MP (Member of Parliament) and MLA (Member of Legislative Assembly). Note that it is possible that a person has become the MP or MLA multiple times so he or she will be at the table multiple times. Write a Query to determine the name of the person who was elected as a MP and also MLA and had the longest time gap between the two elections.	5

		Note: The date column values in the database are stored in YYYY-DD-MM format. (Use Table: - MLA, MP) (5 marks)	
4	a(iii)	 Write a query to display Songname,Artistname,Song Length,Artist Id with longest song length using windows analytical functions. (Use Tables: Song_artiste)(5 Marks) Sample Output songName artisteName songLength artist_id New Fish Thomas Newman 3 1 Rock Hammer Thomas Newman 3 1	5

4	a(iv)	Write a query to create a View on movie title, rating, genre type and runtime. Using the View, Write a Query to fetch the highest runtime movie for each genre type. (Use Tables: Movie, Movie_Studio, Studio, trailer, Genre, Movie_Genre, song, Song_artiste, Artiste) (6 Marks) Sample Output movieTitle movieRating genreType movieRuntime ranking The Dark Knight 5 Action 152 1 Interstellar 5 Adventure 169 1	6
4	a(v)	Write a Query to display the Movie details like id , title , genre and runtime of such movies which have maximum run time and exclude the movie with the name 'Two imprisoned men' from the list. (Use Tables: Movie, Movie_Studio, Studio, trailer, Genre, Movie_Genre, song, Song_artiste, Artiste)(5 marks) Sample Output movieID movieTitle genreType movieRuntime 10 The Wolf of Wall Street Comedy 180	5
4	a(vi)	Write a Query to display movie details like title, ratings, genres, release date along with the earliest release date of the movie in its Genre. (Use Tables: Movie, Movie_Studio, Studio, trailer, Genre, Movie_Genre, song, Song_artiste, Artiste)(5 Marks) Sample Output movieTitle movieRating genreType movieReleaseDate first_release_date_of_genre The Matrix 4 Action 1999-03-31 1999-03-31 The Dark Knight 5 Action 2008-07-18 1999-03-31	5