Database systems
Written exam - July 11, 2022
Draft solution

Question 1 (points 1): Theory question - Select the correct answer

The following relational tables are given (primary keys are underlined, optional attributes are denoted as '*'):

USER (Username, Name, Surname, City)
VIDEO (CodV, Title, Category, Size, Duration, Resolution)
VIEWS (Username, CodV, Date, NumberOfViews)

Given the following SQL query

```
SELECT Title, Date
FROM USER U, VIDEO V, VIEWS W
WHERE Category ="Sport" AND City= "Turin"
AND U.Username = W.Username AND V.CodV = W.CodV
GROUP BY V.CodV, Title, Date
HAVING SUM(NumberOfViews) > 100
```

Choose the option that describes the information selected by the SQL query

The SQL query

A. finds videos in the sport category with more than 100 views by users from Turin
B. finds videos with more than 100 views by users from Turin
C. finds users from Turin who have viewed more than 100 videos
D. finds videos in the sport category with more than 100 daily views by users from Turin
E. all previous answers are correct

Answer D

Question 2 (points 1): Theory question - Motivate the answer

The following relational tables are given (primary keys are underlined, optional attributes are denoted as '*'):

ROOM (Cinema, Room, Seats)

MOVIE (MovieID, Title, Duration)

MOVIE_PROJECTION (Cinema, Room, Date, StartTime, EndTime, MovieID)

Check whether the following instances of the tables are consistent with the table schemas above. Justify the answer.

ROOM

Cinema	Room	Seats
C1	1	100
C1	2	200
C2	1	150

MOVIE

MovieID	Title	Duration
M1	Star Wars II	150
M2	Minions	80
M3	Indiana Jones	90
M1	Star Wars I	120

MOVIE_PROJECTION

Cinema	Room	Date	StartTime	EndTime	MovieID
C1	1	D1	T1	NULL	M1
C2	1	D2	T3	T6	M3
C1	1	D2	T3	T4	M4
C1	3	D3	T5	T2	M1

ANSWER:
MOVIE: two records with same primary key M1
PROJECTION:

- **EndTime equal to NULL;**
- **Violation referential integrity constraint: MovieID=M4 not available in table MOVIE;**
- **Violation referential integrity constraint: Pair (Cinema=C1-3, Room=3) not available in table ROOM**

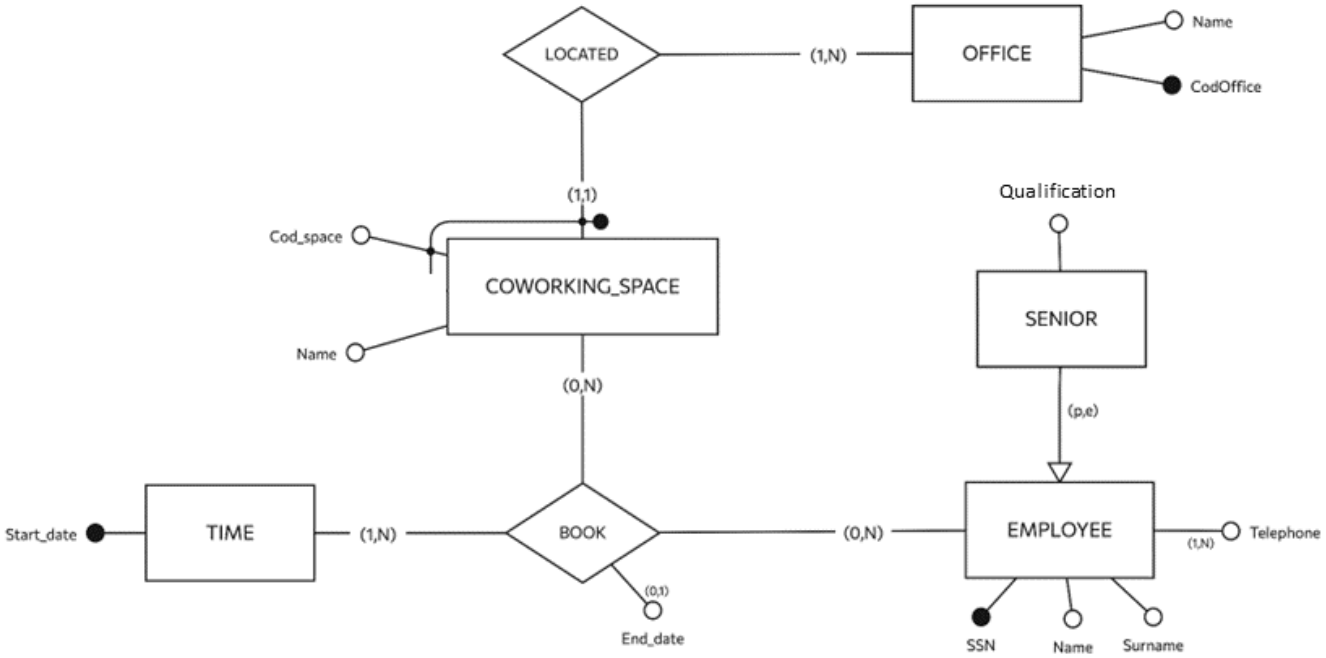
Question 3 (points 1): Theory question - Select the correct answer

In the relational model, it is correct to say that

A. the table schema can include two or more attributes with both the same name and data type
B. attributes in the table schema are identified by means of their position within the schema
C. the table schema cannot include optional attributes (i.e., attributes that can assume the NULL value)
D. none of the answers above is correct

ANSWER D

Given the following conceptual schema represented using the ER diagram



Select the corresponding logical schema (primary keys are underlined, optional attributes are denoted as '*'):

Note: Consider that each *telephone* can be associated with only one employee

ANSWER: B

A	EMPLOYEE (SSN, Name, Surname, Qualification) TELEPHONE (Telephone, SSN) OFFICE (CodOffice, Name) COWORKING_SPACE (Cod_Space, CodOffice, Name) TIME (Start_date) BOOK (Start_date, Cod_Space, SSN, End_date*)
B	EMPLOYEE (SSN, Name, Surname, EmployeeType, Qualification*) TELEPHONE (Telephone, SSN) OFFICE (CodOffice, Name) COWORKING_SPACE (Cod_Space, CodOffice, Name) TIME (Start_date) BOOK (Start_date, Cod_Space, CodOffice, SSN, End_date*)
C	EMPLOYEE (SSN, Name, Surname, Qualification, Telephone) OFFICE (CodOffice, Name) COWORKING_SPACE (Cod_Space, CodOffice, Name) TIME (Start_date) BOOK (Start_date, Cod_Space, CodOffice, SSN, End_date)
D	EMPLOYEE (SSN, Name, Surname) TELEPHONE (Telephone, SSN) OFFICE (CodOffice, Name) COWORKING_SPACE (Cod_Space, CodOffice, Name) BOOK (Start_date, Cod_Space, CodOffice, SSN, End_date)

Question 5 (points 3): Writing queries using the SQL language

Given the following relational tables (primary keys are underlined, optional attributes are denoted as '*'):

EMPLOYEE (SSN, Name, Surname, BirthDate, Nationality)

THEME_PARK (TPCode, ParkName, Address, City, Country)

CAROUSEL (CCode, TPCode, CarouselName, CarouselType)

EMPLOYEE_WORK (SSN, Date, CCode, TPCode)

TICKETS (Date, CCode, TPCode, NumTickets)

Write the following query in the SQL language:

Find name and surname of each employee born after 1982 who has never worked in theme parks located in France.

```
SELECT Name, Surname
FROM EMPLOYEE
WHERE BirthDate >= '1983-01-01'
AND SSN NOT IN (SELECT SSN
                FROM EMPLOYEE_WORK EW, THEME_PARK T
                WHERE EW.TPCode=T.TPCode
                AND Country='France')
```

Question 6 (points 3): Writing queries using the SQL language

Given the following relational tables (primary keys are underlined, optional attributes are denoted as '*'):

EMPLOYEE (SSN, Name, Surname, BirthDate, Nationality)

THEME_PARK (TPCode, ParkName, Address, City, Country)

CAROUSEL (CCode, TPCode, CarouselName, CarouselType)

EMPLOYEE_WORK (SSN, Date, CCode, TPCode)

TICKETS (Date, CCode, TPCode, NumTickets)

Write the following query in SQL language:

Find surname and birth date of each Italian employee who has worked in at least 3 carousels of type aquatic in May 2022.

```
SELECT Surname, BirthDate
FROM EMPLOYEE
WHERE Nationality = 'Italian'
AND SSN IN (SELECT SSN
            FROM EMPLOYEE_WORK EW, CAROUSEL C
            WHERE EW.CCode=C.CCode AND EW.TPCode=C.TPCode
             AND Date>='2022-05-01' AND Date <='2022-05-31'
            AND CarouselType='aquatic'
           GROUP BY SSN
            HAVING COUNT(DISTINCT C.CCode, C.TPCode) >= 3)
```

Question 7 (points 4): Writing queries using the SQL language

Given the following relational tables (primary keys are underlined, optional attributes are denoted as '*'):

EMPLOYEE (SSN, Name, Surname, BirthDate, Nationality)

THEME_PARK (TPCode, ParkName, Address, City, Country)

CAROUSEL (CCode, TPCode, CarouselName, CarouselType)

EMPLOYEE_WORK (SSN, Date, CCode, TPCode)

TICKETS (Date, CCode, TPCode, NumTickets)

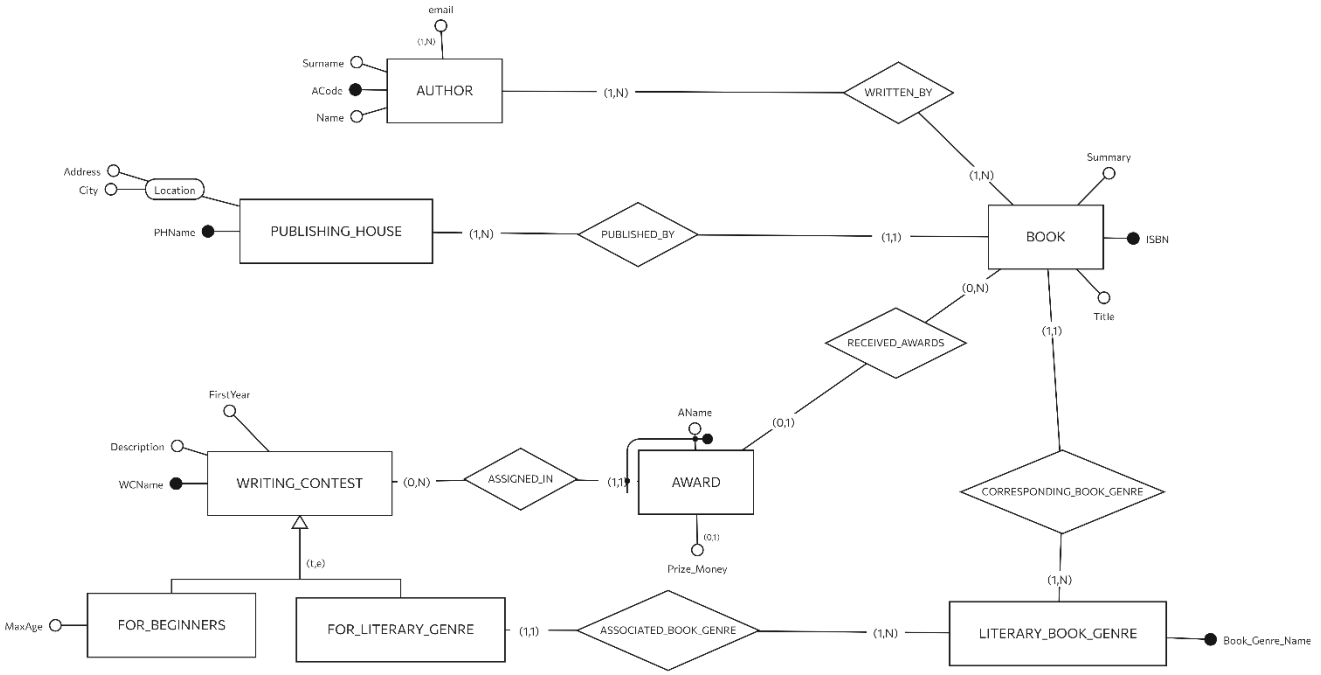
Write the following query in SQL language:
For each theme park that has sold more than 10000 tickets for roller coaster-type carousels, find the name and the city of the theme park and the total number of employees who has worked in the park**.**

```
SELECT ParkName, City, COUNT(DISTINCT SSN)
FROM THEME_PARK T, EMPLOYEE_WORK EW
WHERE T.TPCode=EW.TPCode
AND TPCode IN (SELECT T.TPCode
                FROM TICKETS T, CAROUSEL C
                WHERE T.CCode=C.CCode AND T.TPCode=C.TPCode
                AND CarouselType='roller coaster'
               GROUP BY T.TPCode
               HAVING SUM(NumTickets) >= 10000)
GROUP BY T.TPCode, ParkName, City
```

Question 8 (points 4): Conceptual design of a relational database

You are requested to create a database for the management of some writing contests. Describe the Entity-Relationship diagram addressing the following specifications:

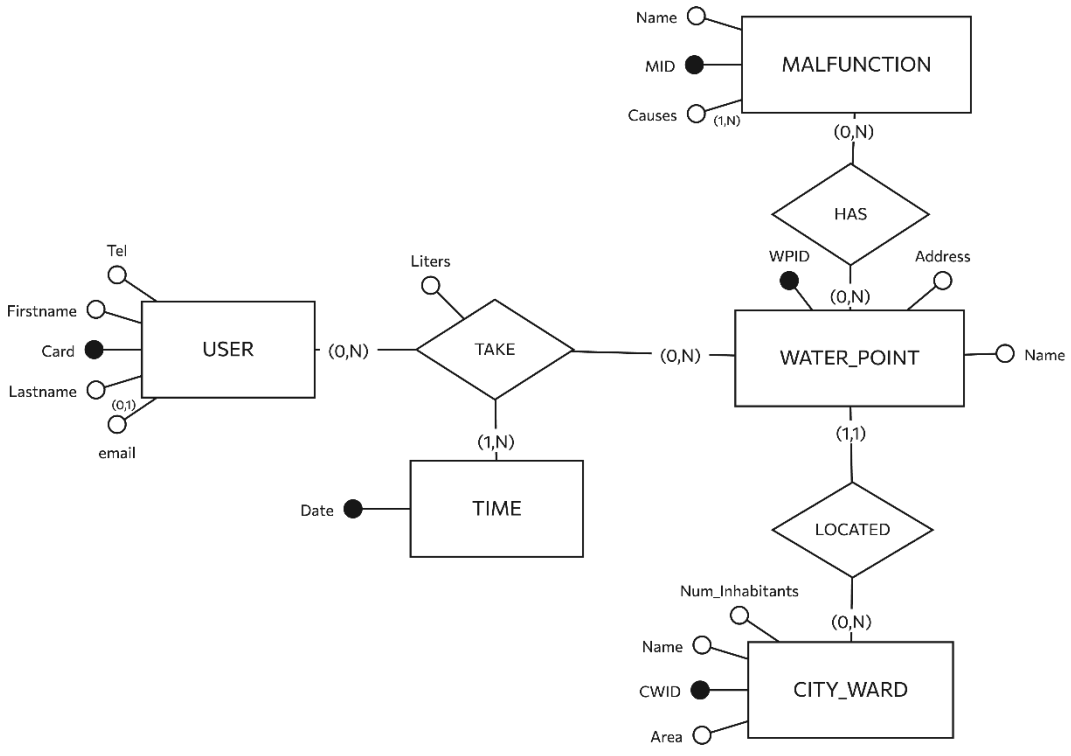
- Each writing contest is identified by a name, and it is characterized by a description and the year the first time the contest was held. The writing contests are of two types: "for beginners" or "for literary genre". For the first type of contest, the maximum age of authors is specified. For the second type of contest, the literary genre that must characterize the book is known.
- Each writing award is identified by means of its name and the writing contest where the award is given. In addition, the amount of prize money, if any, is known.
- Each book is identified by the ISBN code, and it is characterized by the title, a short summary, the literary genre, the publishing house, and the list of authors. Each publishing house is identified by its name; a reference address and city are also known. Each author is identified by a unique numeric code, and he/she is characterized by name, surname and a list of email addresses.
- For each book, the list of writing awards received, if any, is known.



Question 9 (points 5): Conceptual design of a relational database

You are requested to create a database for the management of the “water points” in an Italian city. Water Points are facilities for distributing room-temperature or chilled natural/carbonated water to the public. Describe the Entity-Relationship diagram addressing the following specifications:

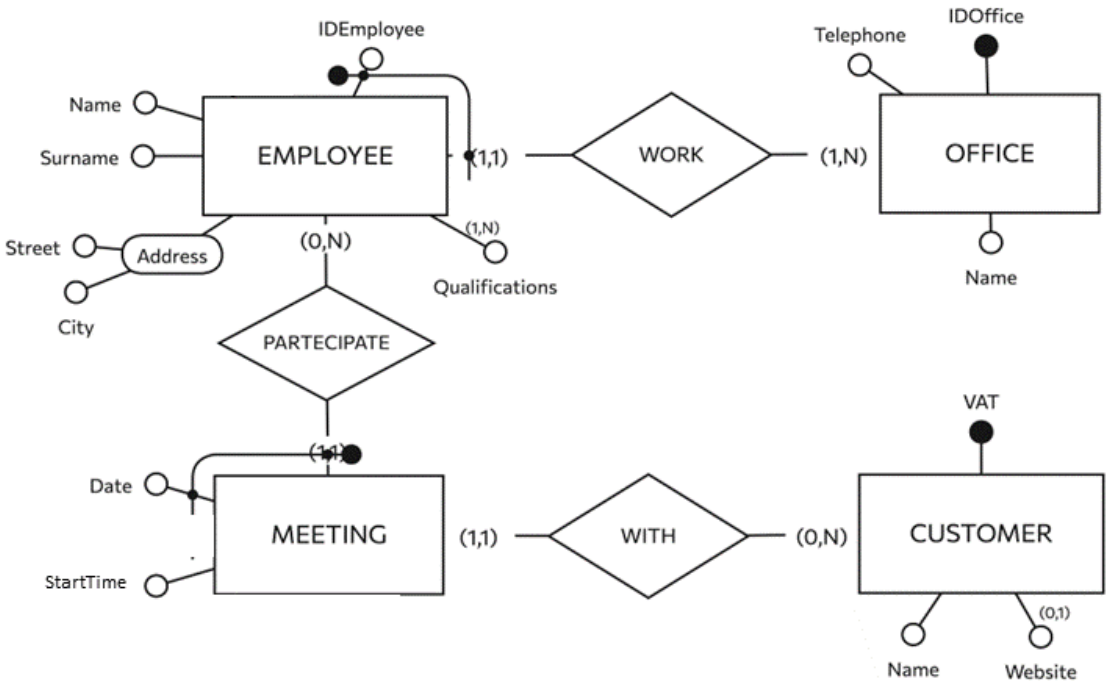
- Citizens can get water through prepaid cards. Each user is identified by the number of his/her prepaid card and characterized by first name, last name, telephone, and e-mail address (if available).
- Each water point is identified by a unique numeric code, and it is characterized by a name. For each water point also the address and the city ward where it is located are available. Each city ward is identified by a unique number. The information on the city ward name, area in square kilometres, and number of inhabitants are also stored in the database.
- You want to store in the database the total amount (in liters) of water taken daily by each user at different water points. Consider that (1) each user can get water from different water points on the same date and (2) different users can get water from the same water point on the same date.
- For each water point the type of malfunctions that have occurred at least once are also stored in the database. Each type of malfunction is identified by a unique code and characterized by a name and a list of possible causes. The same type of malfunction may have been reported for more than one water point.



Question 10 (points 4): Logical Design of a Relational database

Given the following Entity-Relationship diagram, you are required to provide a normalized relational logical schema for the same database.

<u>Note: Consider that each "*Qualification*" can be associated with more than one employee</u>



OFFICE (<u>IDOffice</u>, Telephone)

EMPLOYEE (<u>IDOffice</u>, <u>IDEmployee</u>, Name, Surname, Strees, City)

QUALIFICATION_HELD (<u>IDOffice</u>, <u>IDEmployee</u>, <u>Qualification</u>)

CUSTOMER (<u>VAT</u>, Name, Website*)

MEETING (<u>IDOffice</u>, <u>IDEmployee</u>, <u>Date</u>, StartTime, VAT)

Question 11 (points 5): Data warehouse design

We want to analyse the information relating to the accesses made by users to an online music platform.

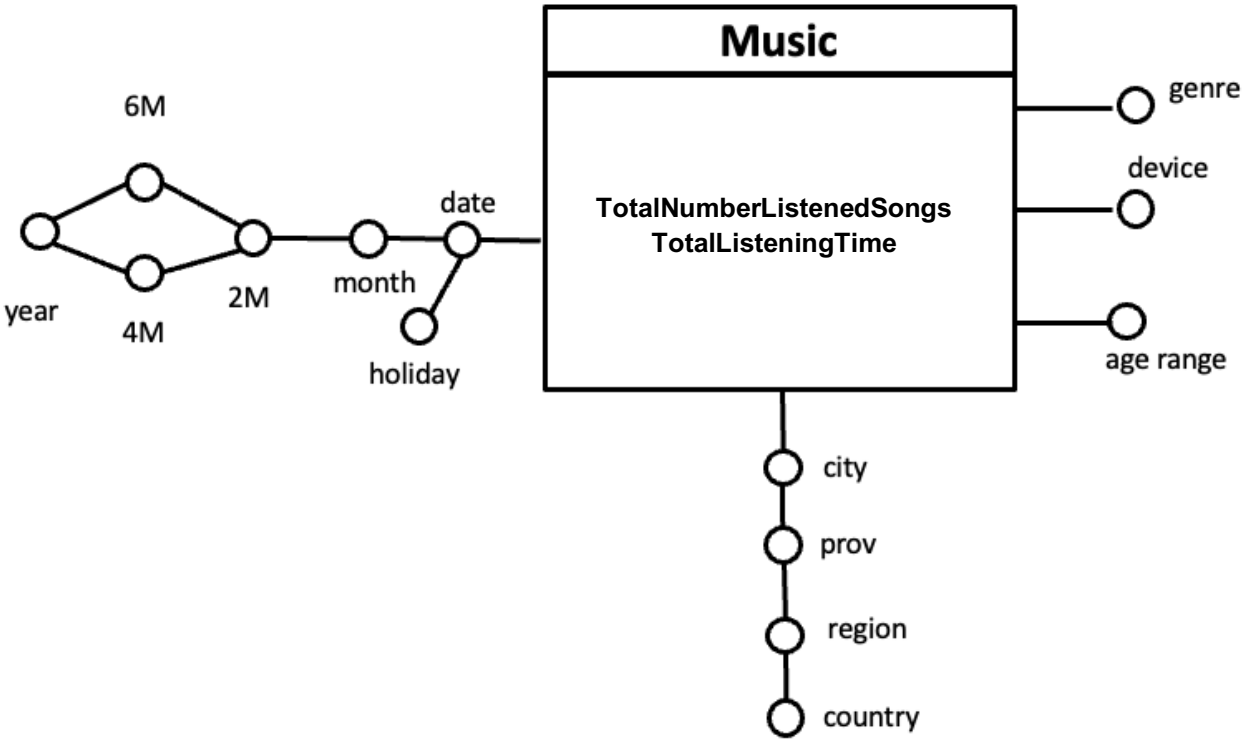
The platform allows users to listen to songs of different genres. The user can access the music content through different types of devices (for example TV or laptop). To guide the user in the selection of the music of his/her interest, each song is described by indicating the genre (e.g., rock or jazz). To support the analysis of interest, the age of the user who accessed the music content is also stored.

We want to analyse the *total number of songs listened to* and *total listening time* for all these songs based on:

- type of device used for listening (e.g., TV, laptop, tablet, or smartphone)
- date, holiday/working day, month, 2-months, 4-months, semester, year
- the content genre (e.g., jazz, rock, or pop)
- user age (<25, between 25 and 40, between 41 and 60, greater than 60)
- city, province, region, country of the user who listened the song

Write the conceptual and logical model of the data warehouse

Conceptual design



Logical design

TIME (TID, date, holiday, month, 2M, 4M, 6M, year)
USER (UID, city, province, region, country)
MUSIC (TID, UID, genre, device, age_range, TotalListeningTime, TotalNumberListenedSongs)