Statistical Methods for Database Integration Examination
DATABASES

The exam consists in two parts:

1) PART A: **The exam is closed-book, closed-notes**;

2) PART B: **You are allowed to use lecture and labs notes**.

Each questions is assigned points expressed in cents.

PART A

Ex. 1

(a) **(10 points)** "The UNESCO Institute for Statistics (UIS) is the official and trusted source of internationally comparable data on education, science, culture and communication. Aiming to evaluate how international trade in cultural goods is relevant for Italian economics, we have selected three main domain for trade in the year **2016**."

Italy	Trade	
2016	**Export**	**Import**
Performance and celebration goods	204.163.335	482.868.210
Visual arts and crafts goods	7.370.364.440	2.385.773.808
Books and press goods	702.612.304	376.969.744

Table 1: Trade per domains in USD

Describe by means of the XML language the data. Specify only one or two domains, but **do not omit to specify all deduced information**.

We need the trade balance value! Is it necessary to describe it?

[*Sol.*:]

```
<cultural-trade>
   <country>
      <name> Italy </name>
      <year> 2016 </year>
      <domain>
         <name> Performance and celebration goods </name>
         <export unit = "USD"> 204163335 </export>
         <import unit = "USD"> 482868210 </import>
      </domain>
   </country>
</cultural-trade>
```

The **balance** could be computed by **export - import**, hence it is not necessary to describe it.

(b) **(10 points)** Assume to transform the table in the point (a) into a single "relation", that could be seen as **a data set**. Write **the relation** specifying for each identified attribute the corresponding **data domain** (refer to the SQL data types).

[**Tip:** *Export* and *Import* header shall be assumed to be values not variables of the data set].

[*Sol.*]

```
Trade(country: VARCHAR(50),
         year: INT,
       domain: VARCHAR(100),
        trade: VARCHAR(10),  [domain = {Import, Export}]
        value: NUMERIC(12,0)
     )
```

(c) **(Optional: 5 points)** *Trip advisor* web site is able to offer you the list of the highest ranked restaurant of the place where you live. Probably to be able to register this data you have to realize a particular process. What the input is? What information you need and what the output is?

[*Sol.*: Considering that we are not the owner of the data, the only way to get data consists in **scraping** the web site. Input is a **HTML** file, we have to capture HTML tags enclosing needed data and the result is a **data set** holding scraped data, typically in the form of a **.csv file** or a **data.frame R object**.]

Es. 2 - Data Modeling

(1) **(35 points)** "Winning Web site should be the tool used to inform people about **accessibility** and **usability** of an historical and artistic beauty, for example: villa (palace), giardino (park), castello (castle)....".

The aim is to design a database to give to interested people historical and practical information by means of a web site.

Draw the E/R diagram that capture the requirements stated below:

(a) We keep track of all accessible **places**, for example the 'belvedere', 'the hunting lodge', 'the powder magazine', ... In addition to the name, we record a description, the maximum number of visitors who can access and its status (open or not in the case of work in progress).

(b) Places information can be extended by means of **one** **card**, shown at the request of the "navigator", with additional information (year of construction, designer, curiosity, and moreover....).

(c) Typically places can be visited in advance virtually, through **multimedia**. All format are allowed: video, picture, sound, ... Obviously a multimedia is strictly related to its place and clearly designed as weak entity set.

(d) Various **tours** are designed for visitors, which touch some or all places. A tour is identified by a color. It could be a guided tour or a free tour.

(e) Places can be made available for private or public **events**, for example weddings, festivals, ... Each event is identified by the type and it has a description. Depending on the relationship between an event and a place there is rule of accessibility (for example, reserved to afternoon events, children are admitted, ...)

(f) To communicate when places are free accessible or they are reserved for tours or they can be booked for events, a **calendar** (calendar days) is planned. Some days will be reserved for free visits and other days for events or for tours.
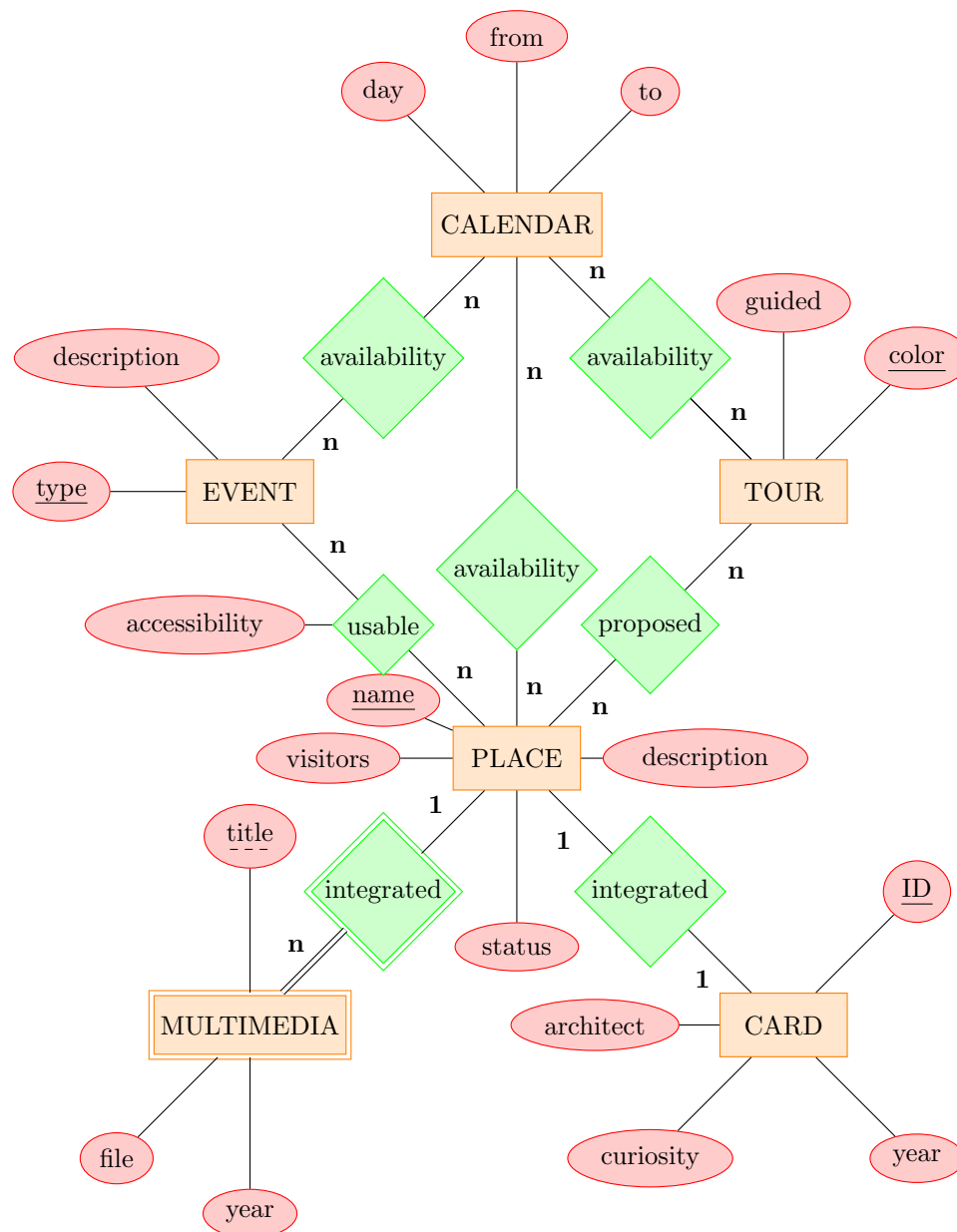
(2) **(Optional: 5 points)**. Write the SQL statement to `CREATE` the "relation" **card**.

```
CREATE TABLE card(
   ID CHAR(6) PRIMARY KEY,
   year INTEGER,
   architect VARCHAR(30),
   curiosity VARCHAR(1000),
   place_name VARCHAR(30),
   FOREIGN KEY (place_name) REFERENCES place(name));
```

PARTE B

Es. 3 - SQL (45 points)   Let assume the database "online-market".

  (1) **Menu(<u>name</u>, description, main)**

  (2) **Food(<u>name</u>, <u>unit</u>, <u>weight</u>, label, price, startDate, endDate, Menu.name)**

  (3) **GiftBasket(<u>name</u>, description)**

  (4) **BasketCombines(GiftBasket.name, Food.name, Food.unit, Food.weight)**

Questions

  1) Create a simple report with for **each gift basket name**, the corresponding **total price**. This price is computed as the sum of the prices of the products the basket combines. Beside the total add a special column with the € symbol for currency.

  [*Sol.*]

```
SELECT basket_name, SUM(food.price) AS price,  'Euro' AS currency
  FROM basketCombines, food
 WHERE basketCombines.food_name = food.name
   AND basketCombines.food_unit = food.unit
   AND basketCombines.food_weight = food.weight
GROUP BY basket_name;
```

  2) We have the need to verify whether in the main-menu **Preserved Foods** there is at least one food having a price larger than 8.50 €. Show the sub-menu they belong to.

  [*Sol.*]

```
SELECT name, main
  FROM menu
 WHERE main = 'Preserved Foods'
   AND EXISTS (SELECT *
                 FROM food
                WHERE price > 8.50
                  AND food.menu_name = menu.name);
```

3) Create a short **price list** including foods of the two sub-menu **Sauces and Pesto** and **Olives and Capers**. For the first sub-menu a special discount of 20% is applied. Report the food name, weight and unit, the price, compute the discounted price and beside a special column with the label **'discounted'** if the price is discounted the label **'full'** if not.

[**Tip:** Select data for each sub-menu separately. For the full prices do not compute anything and just return the NULL value.]

[*Sol.*]

```
(SELECT name, weight, unit, price, price*0.80 AS n_price, 'discounted' AS type
FROM food
WHERE menu_name = 'Sauces and Pesto')
UNION
(SELECT name, weight, unit, price, NULL AS n_price, 'full' AS type
FROM food
WHERE menu_name = 'Olives and Capers')
ORDER BY weight;
```