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)** "Pandemic COVID-19 data is currently mostly analyzed". WHO collected the following data around the world and registered it into a data set. Here below few measures related to Italy.

Date	Code	Country	WHO Region	New Cases	Cum. Cases
2021-04-01	IT	Italy	EURO	23.887	3.584.899
2021-04-02	IT	Italy	EURO	22.184	3.607.083
2021-04-03	IT	Italy	EURO	21.917	3.629.000

Table 1: Pandemic COVID-19: daily regional cases

Describe by means of the XML language the data set. Specify only data of one rows of the table (at your choice). **A country is identified in two different ways. Is it possibile to collapse these data? Is it strictly necessary to describe "cumulative cases"?**

[*Sol.*:]

```
−<covid-19-stat>
   −<day>
        <date> 2021-04-01 </date>
      −<country>
           <code> IT </code>
           <name> Italy </name>
           <region> EURO </region>
        </country>
        <new-cases> 23887 </new-cases>
        <cumulative> 3584899 </cumulative>
      </day>
   </covid-19-stat>
```

Figure 1: **XML file**

A node, in the tree, has data related to country as children. We can omit to describe "cumulative cases" since data represents a time serie and cumulative values could be computed.

(b) **(10 points)** Consider charts created in order to compare quotes of two shares. Assuming to extract useful information, which data would you like to register? What is the corresponding domain?
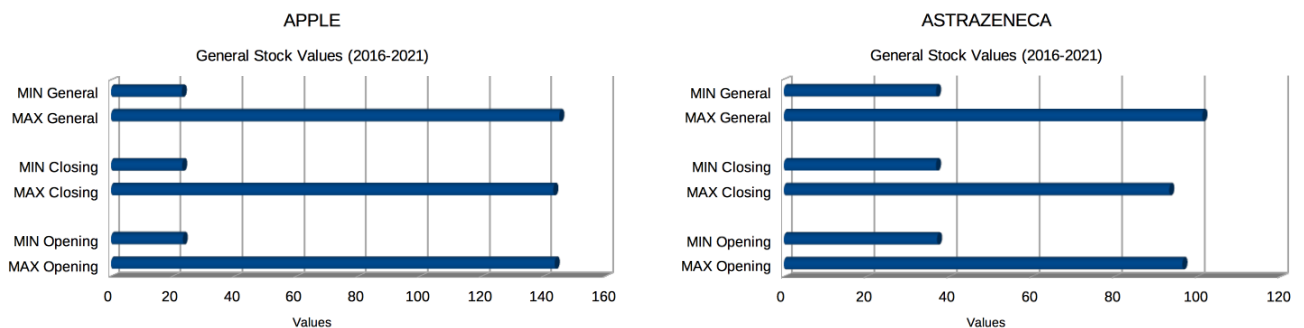
[*Sol.*]

```
Quote(share-name: VARCHAR(30), period: CHAR(9), min-gen: NUMERIC(5,2), max-gen: NUMERIC(5,2),
         min-close: NUMERIC(5,2), max-close: NUMERIC(5,2),
         min-open: NUMERIC(5,2), max-open: NUMERIC(5,2))
```

(c) **(Optional: 5 points)** The main challenge in Data Science is to manage **data**, **information** and **knowledge**. Give a clear definitions of these words.

[*Sol.*: See teaching material]

Es. 2 - Data Modeling

(1) **(35 points)** "Yearly Nations around the world expose and sale their local products at the exposition center of your city. The exposition is mainly dedicated to local manufacturers which use resources offered by the regional agriculture and manufacturing". The goal is to design a database to support an 'app' that promote "virtually" exposed products.

Draw the E/R diagram that capture the requirements stated below. **Use "ID" as key only if strictly necessary**.

(a) At the exposition are admitted only **local products**, realizations of the **agriculture products** (crop, vegetable, milk,...), and local **manufacturing** (wood, cloth, ...). Each product is uniquely identified by a name and material, further it could have a unit and weight, a basic price.

(b) A product comes from a **nation** for which is known and registered the name, continent, capital city.

(c) A product, certified as local realization of a nation, is categorized. A **category** has a code, and a general description. Generally nations are able to expose only some categories of products.

(d) In order to support the 'app' a **picture** of each product is stored. A picture is strictly related to the product which reproduces, therefore **its identifier is strictly related to the identifier of the product**.

(e) It could be an advantage to show not only a single picture for a products but to give the opportunity to browse **details** of the same picture.

(f) Only accredited **manufacturers** can expose. For each ones is registered a code, name and surname or company name, the starting year of its activity, an address, and if available a web site. **A manufacturer is accredited to expose product for a nation and for certain categories**.

(g) In oder to certify the quality of the products the manufacturer should present suitable **certifications**. For each certificate we register the organization who certified and what quality is certified and when.
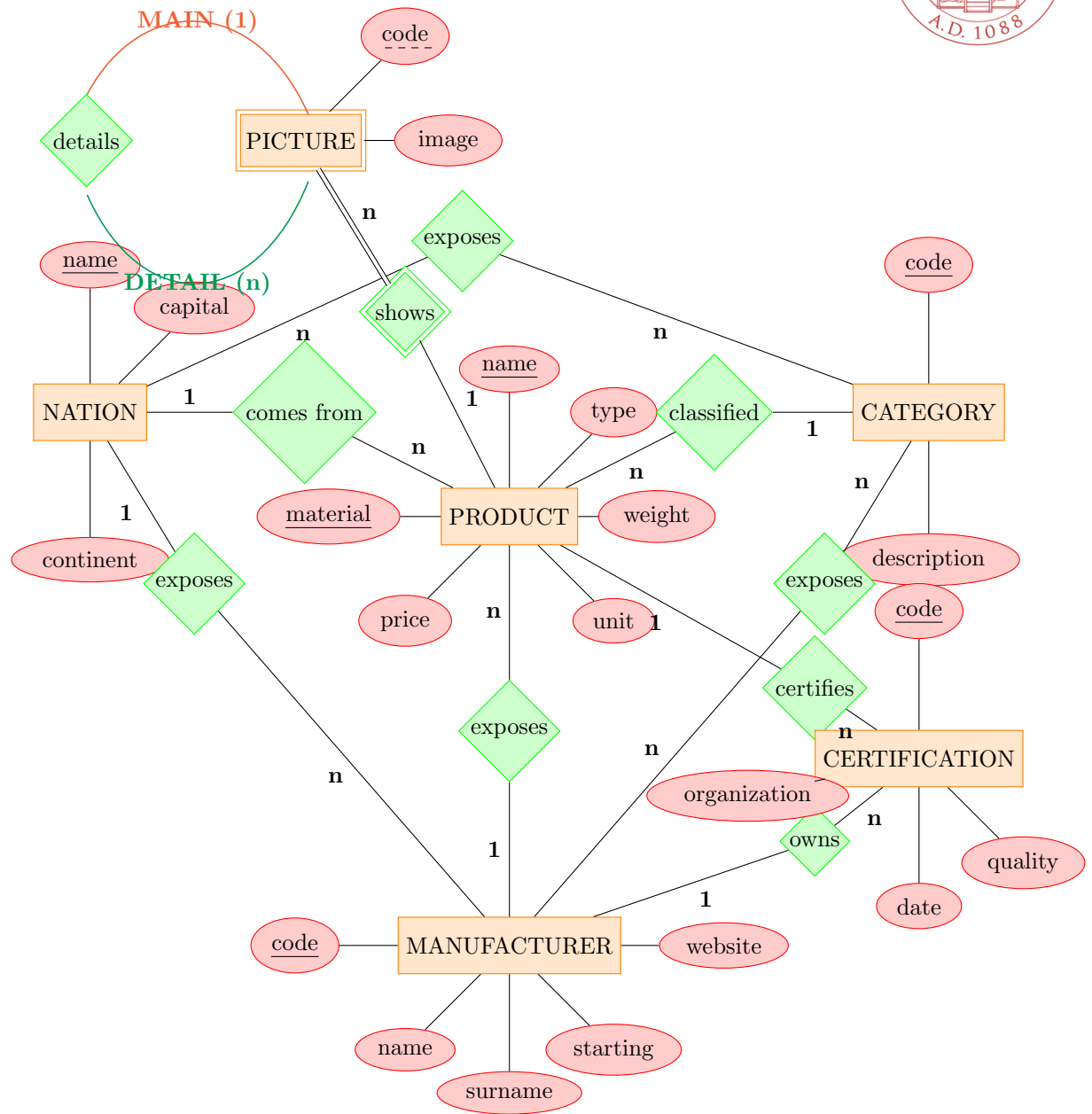
(2) **(Optional: 5 points)**. Write the SQL statement to `CREATE` the "relation" that describes all **certifications** items.

```
CREATE TABLE Certification(
    organization VARCHAR(100),
    date DATE,
    quality VARCHAR(30),
    manufacturer CHAR(10),
    product_name VARCHAR(50),
    product_material VARCHAR(50),
    FOREIGN KEY (manufacturer) REFERENCES Manufacturer(code),
    FOREIGN KEY (product_name, product_material) REFERENCES Product(name, material)
);
```

PARTE B

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

  (1) Region(<u>name</u>, description)

  (2) Sheet(<u>ID</u>, description, Region.name)

  (3) Producer(<u>name</u>, description)

  (4) Produced(<u>Producer.name, Sheet.ID</u>)

  (5) Ingredient(<u>name</u>, description)

  (6) Made(<u>Ingredient.name, Sheet.ID</u>)

  (7) Menu(<u>name</u>, description, <u>main</u>)

  (8) Food(<u>name</u>, <u>unit</u>, <u>weight</u>, label, price, startDate, endDate, Menu.main_name, Sheet.ID)

  (9) GiftBasket(<u>name</u>, description)

(10) BasketCombines(<u>GiftBasket.name, Food.name, Food.unit, Food.weight</u>)

(11) User(<u>ID</u>, date, time, network_info)

(12) Consulted(<u>User.ID, Food.name, Food.unit, Food.weight</u>, time)

(13) Selected(<u>User.ID, Food.name, Food.unit, Food.weight</u>, time, quantity)

Questions

  1) Marketing manger would like to rename gift baskets emphasizing products "value", therefore he would like a report, as below. [**Tip:** apply the strategy to find maximum and minimum separately]

Basket Name	Price	Value
Delight taste	4.60	MIN
Delight taste	8.98	MAX
...	...	...

Use the explicit JOIN whenever it is necessary to join tables.

[Sol.]

```
(SELECT B.basket_name, MAX(F.price) AS price, 'MAX' AS value
   FROM food F JOIN basketCombines B
     ON (F.name = B.food_name
    AND F.unit = B.food_unit
    AND F.weight = B.food_weight)
GROUP BY B.basket_name)
UNION
(SELECT B.basket_name, MIN(F.price) AS price, 'MIN' AS value
   FROM food F JOIN basketCombines B
     ON (F.name = B.food_name
    AND F.unit = B.food_unit
    AND F.weight = B.food_weight)
GROUP BY B.basket_name)
ORDER BY basket_name, price;
```

2) Are there available [currently: endDate `IS NULL`] at least two distinct products (at least in name and price) of the same manufacturer for the same menu? [**Tip**: identify manufacturer by means of `food.label`]

[Sol.]

```
SELECT DISTINCT F1.label, F1.menu_name
  FROM food F1, food F2
 WHERE F1.label = F2.label
   AND F1.menu_name = F2.menu_name
   AND F1.price > F2.price
   AND F1.name <> F2.name
   AND (F1.endDate IS NULL AND F2.endDate IS NULL);
```

3) It could be helpful to analyze food product selections in specific time intervals. Are there **days** [= date] six months ago, such date users selected at least 3 products? [**Tip:** consider the process to count number of selections six months ago]

   [*Sol.*]

```
SELECT date, COUNT(User.ID) AS nr
   FROM User JOIN selected ON User.ID = selected.ID
WHERE month(date) = month(now()) - 6
GROUP BY date
HAVING nr > 3;
```