



# 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)** “Online shops are mostly used to buy cloths”. Consider the promotional advertisement below of an online shop for kids. Report data in a semi-structured data model (XML Language). **Describe at least two cloths!**

[**Tip:** the small numbers at the bottom correspond to cloth sizes.]

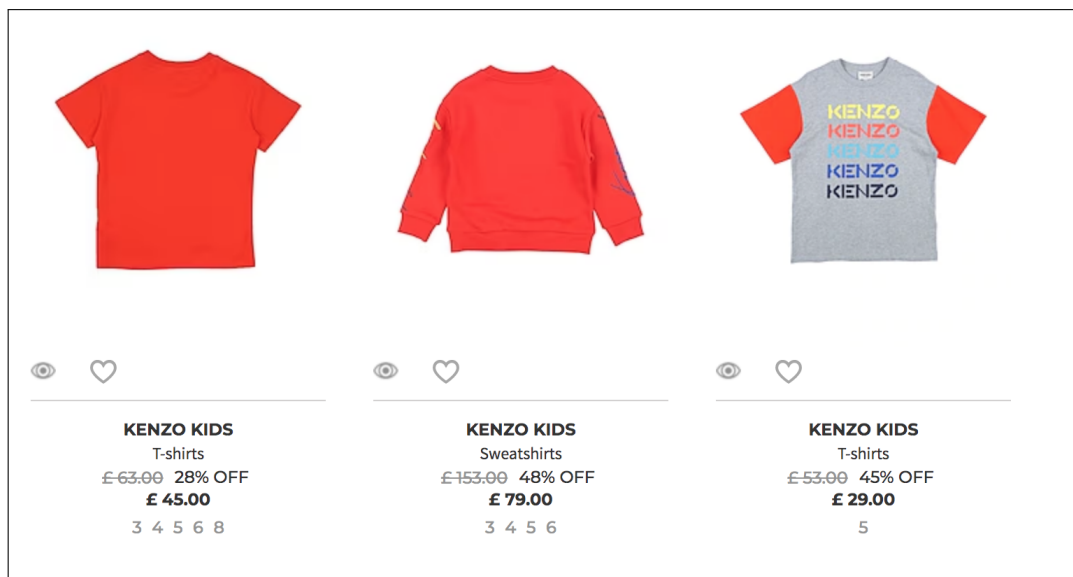


Figure 1: source: YOOX Online Shop



```
<shop-kids>
  <brand>
    <name> Kenzo Kids </name>
    <cloths>
      <item>
        <type> Sweatshirts </type>
        <price unit="£"> 153.00 </price>
        <save unit="%"> 48 </save>
        <sizes>
          <s1> 3 </s1>
          <s2> 4 </s2>
          <s3> 5 </s3>
          <s4> 6 </s4>
        </sizes>
      </item>
      <item>
        <type> T-shirts </type>
        <price unit="£"> 53.00 </price>
        <save unit="%"> 45 </save>
        <sizes>
          <s1> 5 </s1>
        </sizes>
      </item>
    </cloths>
  </brand>
</shop-kids>
```



- (b) **(10 points)** Given the relation that aims to hold data about kid cloths, create by means of the SQL statement the corresponding table. Further write two statements in order to insert data, at your choice!

Cloth(genre: {man, woman, kid}, type, date, price,  
save: {yes, no}, percentage {default 0},  
appreciation: {high, medium, low})

In braces are listed the acceptable values.

```
CREATE TABLE Cloth(  
    genre VARCHAR(5),  
    type VARCHAR(50),  
    date DATE,  
    price NUMERIC(5,2),  
    save BOOLEAN,  
    percentage NUMERIC(4,2) DEFAULT 0.00,  
    appreciation VARCHAR(6),  
    ID CHAR(6) PRIMARY KEY  
);  
  
INSERT INTO Cloth VALUES('kid', 'T-Shirts', '2023-02-02',  
    153.00, TRUE, 48.0, 'medium', 'I00001');  
  
INSERT INTO Cloth VALUES('man', 'T-Shirts', '2023-02-02',  
    95.80, TRUE, 38.0, 'high', 'I00002');
```

- (c) **(Optional: 5 points)** What does it mean ‘Normalization’? What does BCNF say?

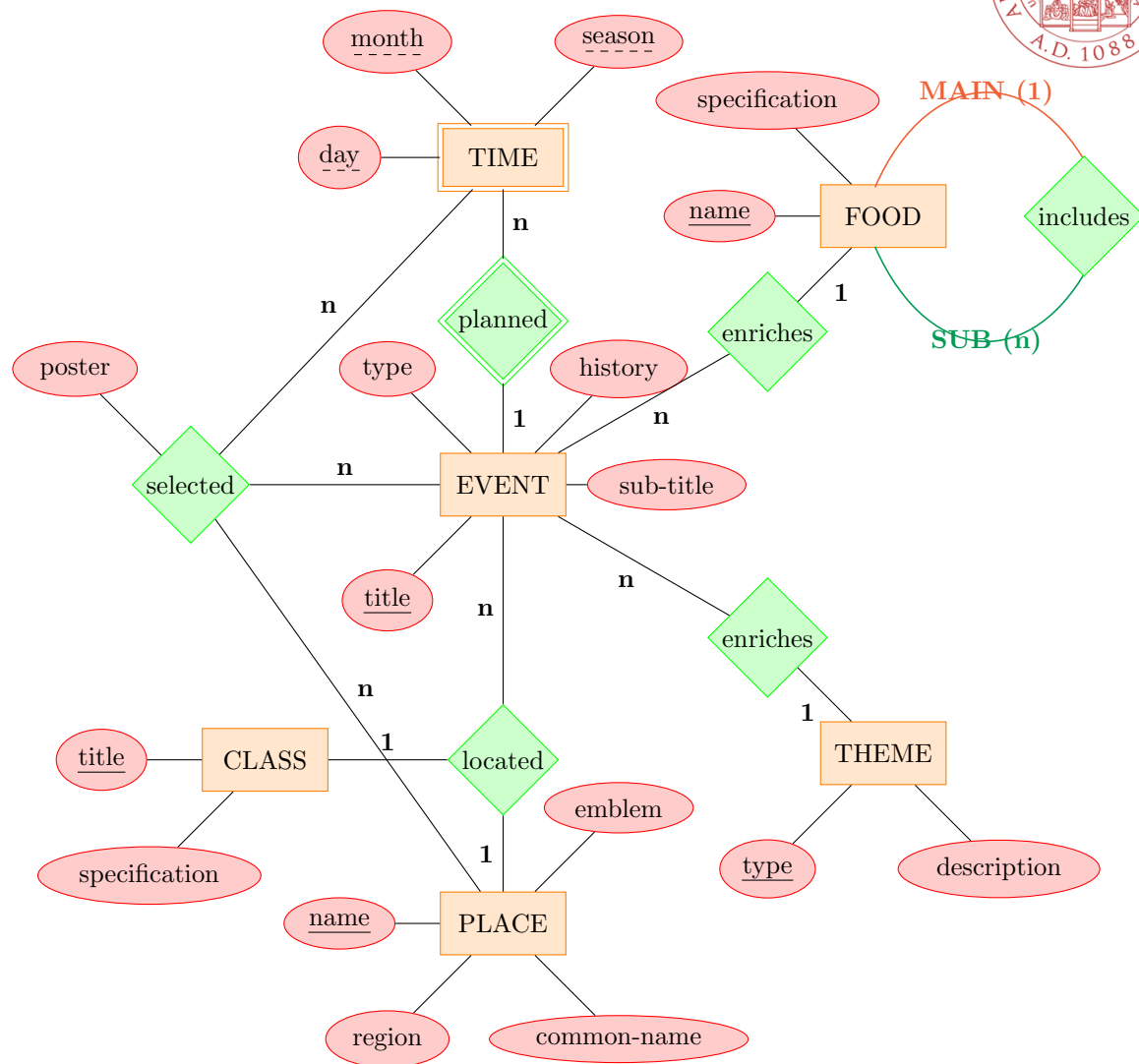
[Sol.: See teaching material]

**Es. 2 - Data Modeling**

- (1) (35 points) “enjoy@fair **application** advises tourist who likes enjoying in the spare time at the traditional fairs and festivals.” Thanks to the functionalities offered the user browses all events which could be regional fairs or festivals, even to consider selections by theme, season ... area. The goal is to design a database supporting the ‘app’.

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

- (a) All available information on fairs and festivals is collected and registered as **event**, specifically for each one type, title and in case a sub-title, mention about their history. Considering that an event is either a fair or a festival or both, and that the related information could vary we aim to distinguish it.
- (b) Firstly **food**, with its name and specification, which generally specifies fairs (for example jam fair, truffle fair, pork fair, ....). *In few cases we have that many foods would be registered with a generic name, for example pork, which includes, ham, salami, .. this fact has to be captured.* Then **theme**, with type, for example harvest, patron saint,.. and at least a description. These registrations (both) should integrate information of *event*.
- (c) Event has to be integrated with more information. **Time** registers information about when the event holds, specifically the season, month and day. An event could be held many times in one season or year. Identify a proper and informative key.
- (d) Place completes the set of information. **Place** holds name, common name, region, and emblem of the village/city.
- (e) Event and place information is sufficient to assign a **class** to the event (for example Bologna *antiques* fair, Reggio *farm* festival, ...). Class holds title and specification.
- (f) Selections of events are generally defined, creating a *poster* showing the list of events, with places and time they holds. The poster, which is a document, is registered.



- (2) (Optional: 5 points). Write “relations” of the **Relational Data Model** which describe **Event** and **Time**.

```
Event(title, sub-title, type, history,
      place.name, class.title, theme.type, food.name)
Time(day, month, season, Event.title)
```



## PARTE B

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

- (1) Region(name, description)
- (2) Producer(name, description)
- (3) Sheet(ID, description, Region.name, Producer.name)
- (4) Ingredient(name, description)
- (5) Made(Ingredient.name, Sheet.ID)
- (6) Menu(name, description, main)
- (7) Food(name, unit, weight, label, price, startDate, endDate, Menu.menu\_name, Sheet.ID)
- (8) GiftBasket(name, description)
- (9) BasketCombines(GiftBasket.name, Food.name, Food.unit, Food.weight)
- (10) User(ID, date, time, network\_info)
- (11) Consulted(User.ID, Food.name, Food.unit, Food.weight, time)
- (12) Selected(User.ID, Food.name, Food.unit, Food.weight, time, quantity)

### Questions

- 1) We would like to create a particular gift basket having in mind something of salad and something of sweet excellence. List name, weight, unit of foods of something different from ‘Sweet Products’ and ‘Preserved Foods’ (**main menus**). Add foods which report in their description (**published on the sheet**) the name of the two excellences: ‘gianduia’ or ‘pistacchio’. Add the information ‘SALAD’ or ‘SWEET’ at the end of each row.

```
(SELECT food.name, food.weight, food.unit, 'SALAD'
  FROM food, menu
 WHERE food.menu_name = menu.name
       AND main NOT IN('Sweet Products', 'Preserved Foods'))
UNION
(SELECT food.name, food.weight, food.unit, 'SWEET'
  FROM food, sheet
 WHERE food.sheet_ID = sheet.ID
       AND (description LIKE '%gianduia%' OR description LIKE '%pistacchio%'));
```



- 2) The report of the last year showed that on January and on July users accessed to the online market the most. Marketing asks us to analyze user accesses on **January** and on **July** showing users (**identify a user by user network\_info**) who accessed **at least twice**.

```
SELECT DISTINCT month(u1.date), u1.network_info
FROM user u1, user u2
WHERE u1.network_info = u2.network_info
AND u1.date < u2.date
AND month(u1.date) IN (1,7)
ORDER BY u1.date;
```

- 3) The statistics of food appreciation is necessary to make suitable sale strategies, therefore:

**A)** Compute for each menu and for each producer description (**in attribute 'label'**), the number of selected food items (in the basket to buy), including those producers whose food items have never been selected.

**B)** [extra points] Reports only data with the best statistic, that is **the largest number of selected items**.

```
SELECT menu_name, label, count(selected.food_name) AS nr
FROM food LEFT JOIN selected ON
    (food.name = selected.food_name
    AND food.unit = selected.food_unit
    AND food.weight = selected.food_weight)
GROUP BY menu_name, label
HAVING nr >= ALL (SELECT count(selected.food_name) AS nr
                  FROM food LEFT JOIN selected ON
                      (food.name = selected.food_name
                      AND food.unit = selected.food_unit
                      AND food.weight = selected.food_weight)
                  GROUP BY menu_name, label);
```