Databases
Session 3

1. OBJECTIVE:

1. Understand what is an entity, know how to identify its attributes, understand the role of the PK (and the CKs) and the relational structures (tables) from 2 use cases.
2. Experiment with client-server architecture.
3. Become familiar with relational DB software (Oracle): DataModeller (design) and SQLDeveloper (creation, manipulation, and querying)

2. AUTONOMOUS WORK

Previous Preparation (Recommended): Review and read the documents Session3_DesignIntroduction.pdf and for Session3_BasicERDesign1.pdf.

3. MATERIALS

Theory: Session3_DesignIntroduction.pdf, Session3_BasicERDDesign1.pdf.
Problems and Practices: DataModelerTutorialBasic_EN.pdf

4. EXERCISES

Requirement: Supermarket

A supermarket wants to create a database of sales made to ATMs. This database must keep information about the tickets and the products available in stock. For each product we want to save the name, brand, price, units in stock and barcode. The barcode is an alphanumeric code that has 5 numbers and a letter at the end. For each sale, the ticket that must contain the total amount, date (day and time), form of payment (cash or card) and box where the purchase was made. We also want to know how many units were sold for each product

Exercise 4.1. Identify entities and their attributes.
Attributes for products: name, brand, price, units in stock, units sold, and barcode (5 numbers and letters).
Attributes for tickets: total amount, date (day and time), form of payment (cash or card), box.
Entities: Tickets and products

Exercise 4.2. Identify candidate keys (CK) and select a primary key (PK). Explain What the attributes have to fulfill to be CK and identify sets of attributes that cannot be CKs.
CK: barcode, date, box.
PK: barcode, date and box

Exercise 4.3. Open the DataModeller and create the entities with the CK chosen in Exercise 4.2. Assume that all attributes are simple.

Exercise 4.4. Translate the previous design to a Relational Model. What structures were created? Save your work for use in the other sessions.

Exercise 4.5. Generate the data structure SQL script with the DataModeller. Open SQL Developer and execute the script. What does it take for the script to run? Why? Relate it to any of the BD architectures.

Exercise 4.6. Insert the following data:

> Requirements: (3, 30/09/18, card, 3€), (3, 01/10/18, card, 3.25€), (01, 30/09/18, cash, 3.4€), (Milk, Vacona, 0.75, 50, 56790A), (Milk, La munyidora, 1.25, 70,12345Z), (Beer, El monjo, 2.15, 100, 54389Q)

Use the table data tab or the command line insert statement:
Insert into product (name, brand, price, units, barcode) values ('Beer', 'El monjo', 2.15, 100, '54389Q');

Exercise 4.7. Now, try to insert this data:

> Requirements : (Milk, Vacona, 1.25, 70, 56790A),

Can you do it? Why?

Exercise 4.8. Use DataModeller to create the tables without defining the CK, insert the Exercise 4.6 data, and then do Exercise 4.7 again. What basic property of DBSs is being breached?

5. EXERCISE FOR HOMEWORK

Requirement: Artificial Intelligence

UAB researchers in collaboration with other universities are developing an AI system for the early detection of breast cancer using radiological measurements. In order to develop the diagnostic support system, they have to test different pattern classification algorithms in a set of cases already labelled with their final diagnosis. Researchers need to save the results of those experiments into a DB so that other universities in the project have access to them.

In each case they want to save the case ID (the data have been anonymized according to the data protection regulations), the radiological measurements (average intensity of the lesion, size, eccentricity and kurtosis of the histogram) and the final diagnosis (inflammatory, ductal carcinoma, lobular carcinoma, papillary carcinoma). Of each classification method, they want to save the name, researchers who developed it, and the values of the parameters with which it was executed. Each classifier is tested in a different set of cases chosen at random. They want to save what these cases are and if the method correctly predicted the final diagnosis.

Exercise 1. Identify entities and their attributes.

Entities: Case and classification methods

Attributes for cases:

- Case_ID
- Radiological_measurements
- Lesion_intensity_average
- Size
- Eccentricity
- Kurtosis_histogram
- Final_diagnosis

Attributes for classification methods:

- Algorithm_name
- Researchers
- Parameters_values

Attributes for the set of cases with the info telling if the algorithm prediction was correct:

- Random_case
- Correct_prediction: as a boolean (True/False)

It might be also useful to have another table with the relations between each table so that we can easily navigate through all the info without any problems.

Exercise 2. Identify candidate keys (CK) and select a primary key (PK). Explain why you choose each attribute as key.

A candidate key is an attribute that can be used to identify each case in a table.
For the 1st table the only possible CK is the case ID because it is the only one that will be 100% different for each case. The other ones are parameters that will depend on measurements and the diagnosis, which is possible for them to have repeated values along different cases making them not suitable to be identifiers.
For the 2nd table there is also one possible CK as long as one developer can create more than one algorithm and the parameters can be repeated along different executions of the algorithm. Each algorithm is very likely to have a different name making the name of the algorithm a suitable CK.
In the 3rd table there is one possible case too. The randomly chosen cases can be used as a CK because each one will have a different value making it easy to identify each case. The correct prediction attribute only tells us if a prediction is correct or not, thus is not suitable for a CK.

CK: case ID, name, random case

There is only one CK for each table so they will be our PK
PK: case ID, name, random case