Sistemas de Informação e Bases de Dados

1º semestre

Exame de 1ª Época

This exam is written in English, but you may answer in Portuguese.

Part I: E-R modeling **Answer in a separate sheet of paper**

1. You have been hired to design a database for an airline company. For this purpose, you are given the following description:

   Each flight has a flight number and a certain date and time (example: AF1025 on January 12, 2015, at 11:55). There may be the same flight number in different dates. The flight departs from an airport and arrives at another airport; in some cases, there may be an intermediate stop at a third airport. Airports are identified by an airport code, but they also have a name, a city, and a country (example: "CDG" is for Charles de Gaulle International Airport, in Paris, France). A flight uses a certain aircraft. An aircraft has a manufacturer, a model, and a capacity (example: Airbus A319 has 124 seats). Each flight has a list of passengers. For security reasons, we need to know about each passenger: id, first name, last name, birthday, nationality, and country of residence. In each flight, each passenger travels in a certain class (F = first class; J = business class; Y = economy class). A passenger may have one or more pieces of luggage. Each piece of luggage has a barcode and weight.

   (a) Draw an E-R diagram for this scenario. Include appropriate constraints in each association.
   (b) Convert your E-R model into a set of tables. Indicate the primary and foreign keys in each table.

Part II: SQL **Answer in a separate sheet of paper**

2. You arrive at the airport and you see the following information about arrivals and departures. Think of this information as being two database tables.

Arrivals

Time	Flight	Origin	Status
11:25	LH1166	Frankfurt	Arrived
11:30	TP1043	Barcelona	Arrived
11:45	KL1693	Amsterdam	Scheduled
12:00	EK191	Dubai	Scheduled

Departures

Time	Flight	Destination	Status
11:30	EZY7637	Bordeaux	Departed
11:30	TK1756	Istanbul	Departed
11:35	TP838	Rome	Last call
11:55	AF1025	Paris	Boarding

   Write a SQL query to answer each of the following questions:
   (a) Which cities is TAP Portugal flying to or from? (Consider both arrival and departure flights.) Note that TAP flights are identified by a flight number beginning with "TP".
   (b) Suppose that a passenger arrives from a city and needs to go back to the same city. Which cities appear both in the arrivals and in the departures, with the departure time being later than the arrival time?
   (c) If the same flight number appears both in the arrivals and the departures, it means that the flight is making an intermediate stop at this airport. List all flights that are making an intermediate stop at this airport, together with the duration of the intermediate stop.

(d) Which destination city has the largest number of departing flights? Consider only the flights that are in the "Departed" state.
(e) Find which cities appear either in arrivals or in departures, but not in both.

3. Consider the same information about arrivals and departures as above.
   (a) Write a SQL function called **num_arrivals**(…) to find how many flights are arriving from a given city. The city name is passed as an input parameter to the function.
   (b) Imagine that you have a similar function called **num_departures**(…) to find how many flights are departing to a given city. Using both functions, write an SQL query to determine the city which has the largest difference between the number of arrivals and the number of departures.

Part III: HTML and PHP **Answer in a separate sheet of paper**

4. Suppose you want to develop a Web-based application that allows a user to check if a given flight has already arrived. In the database, you will have to run a query similar to the following:
   SELECT Status FROM Arrivals WHERE Flight = 'LH1166' ;
   where the flight number is given by the user.
   (a) Write a static HTML form that allows the user to submit a flight number in a text box.
   (b) Write a PHP script to get the flight number submitted in the form, execute the SQL query, and show the result to the user.

Part IV: Advanced Topics **Answer in a separate sheet of paper**

5. To simplify things a bit, we will assume that flight numbers contain only digits. Suppose that we create an index with the following flight numbers: 1166, 1043, 1693, 191, 7637, 1756, 838, 1025 (in this order). The index is a B^+-tree with a maximum of two values in each node. Draw the B^+-tree after the insertion of each value.

6. Suppose that you have 2 airports, and each airport has its own database of arrivals and departures. You run the following query in both databases:
   EXPLAIN SELECT * FROM arrivals AS a, departures AS d WHERE a.origin = d.destination;
   In the first database you obtain an execution plan with a hash-join, and in the second database you obtain an execution plan with a merge-join. What could be the explanation for this? Justify your answer and describe your knowledge about these algorithms.

7. The ground staff at the airport is on strike, and your boss gives you a list of flights to be cancelled. You need to update the status of those flights to "Cancelled", in a single transaction. However, other people are updating the table at the same time (changing the status of other flights, adding new flights, etc.). You remember having learned something about "isolation levels" at the university. Describe your knowledge about isolation levels and explain how you can use them to allow multiple people to work on the table at the same time.

Sistemas de Informação e Bases de Dados

2016-01-15

Exam

You may answer the questions in English or Portuguese. Duration: 90 minutes

Part I: E-R modeling

You have been hired to design a database for a manufacturer of electronic devices. For this purpose, you are given the following description:

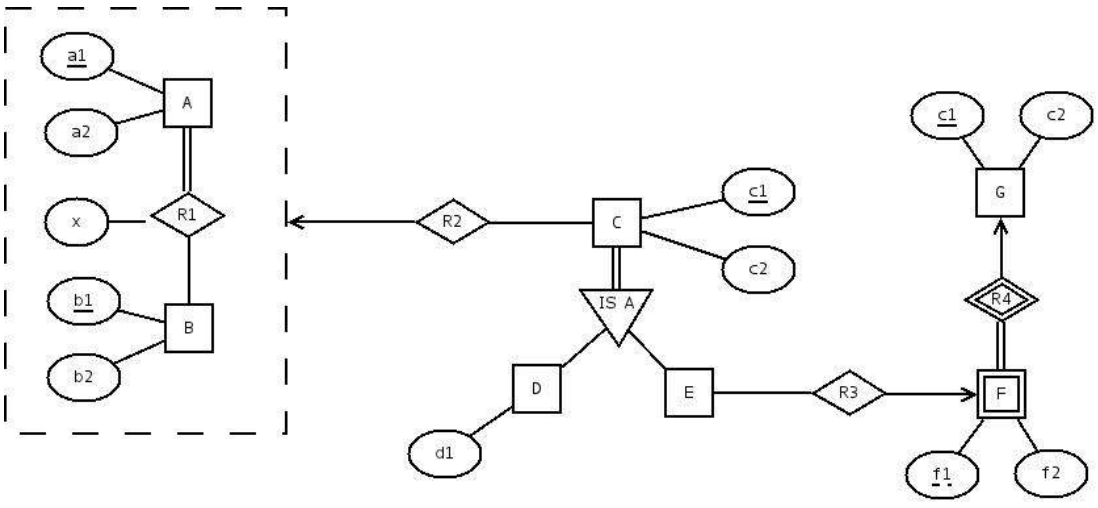
Chip Electronics is a manufacturer of wearable electronic medical devices. Each device has a serial number, a description and a list of components. Components are procured from suppliers; for management purposes, each component must have at least one supplier, which has a name, an address, and a phone number. For each component, the name and a description are stored in the database. The manufacturer must also keep track of the components in stock. For each component, there may be several units in stock, and each unit has a number and a procurement date. Units from the same component have distinct numbers, but units from different components may have the same number. After the devices are assembled, the manufacturer sells the medical devices to medical centers for a given price and quantity, where each center has a name and an address.

Answer the following questions in a separate sheet of paper

1. [2.5 pts] Draw an E-R diagram for this scenario. Include appropriate constraints in each association.

2. [2.0 pts] The following requirements have to be included in the database design:
   The medical devices are now sold to medical centers through two different modes. The first mode uses a long-term contract, in which the medical center buys a fixed number of devices per month for a given price and for a fixed period of time. The second mode, to procure components for specific needs, uses one-off contracts, which have a price and a purchase date. In addition, all contracts have an identification number and a description.
   Update the previous E-R diagram in order to incorporate the above requirements.

3. [1.5 pts] Convert the following E-R diagram to the relational model.



Part II: SQL and PHP

Consider a database for managing energy consumption in a large organization. There are energy meters scattered across several campi that record the count of energy units consumed (Kwh) at the end of every hour. There are 2 energy rates, A and B, for two periods of consecutive hours.

Meter

MeterID	Building	Department	Campus
001	NorthTower	EE	Alameda
002	NorthTower	EE	Alameda
003	SouthTower	BioEng	Alameda
101	ISTTagus	EE	Taguspark
202	InformaticaII	CSE	Alameda

Rate

RateClass	UnitPrice	StartHour	EndHour
A	0.20	08	22
B	0.10	23	07

Reading

MeterID	Month	Day	Hour	Units
001	12	25	00	1023
002	12	25	00	255
001	12	25	01	1024

Answer the following questions in a separate sheet of paper

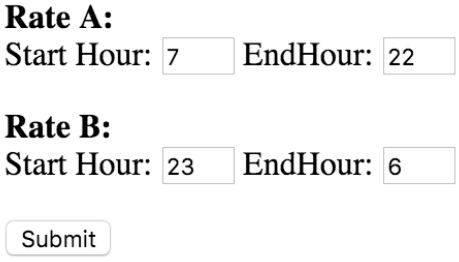
4. [4.0 pts] Write an SQL query for each of the following questions:
   a) How many meters has the EE department which are not installed in the NorthTower?
   b) How many readings have been collected at consumption Rate A in Taguspark?
   c) Which building had more readings in Month 12?
   d) Which buildings have installed meters for all departments?

5. [1.5 pts] Write an SQL function, *last_reading(meter, month),* that returns the last recorded measure (units) of a given meter in a given month. (Hint: since units are always increasing, the last reading is equal to the maximum reading for that meter/month.)

6. [1.0 pts] The energy consumption measured by meter M in month X is given by *last_reading(M, X) – last_reading(M, X-1)*.
   Write a query to obtain the total energy consumption in month 8 (considering all meters) using the function above.

7. [1.0 pts] Write the HTML form to set the rate classes (content of table "Rates" in previous question). The form should have the layout of the figure and invoke the URL:

   http://host.pt/setRates.php?
   startA=7?endA=22?startB=23?endB=6



8. [1.5 pts] Write the PHP code of *setRates.php* to update the two rows of table "Rates", for the rate "A" and rate "B" when run by a Web server upon receiving the URL of the previous question.

Part III: Advanced topics

Answer the following questions in a separate sheet of paper

9. [2.0 pts] Consider the following PHP instruction:

```
$result = $connection->query("SELECT balance FROM account WHERE
account_number='{$_REQUEST[account_number]}'");
```

   (a) Is this instruction vulnerable to an SQL injection attack? If so, explain how the attack could be carried out. If not, justify why not.
   (b) Explain how you can protect against such attack. If you remember the code, present it. If not, describe it in words.

10. [1.5 pts] Suppose that there are two transactions (T1 and T2) running in parallel in READ COMMITTED mode, as shown in the figure. T1 updates the value of some object A (e.g. the balance of an account). T2 reads the value of A at three different points in time.

T1	T2
	select A
update A	
	select A
commit	
	select A

   (a) At each "select A" instruction, does T2 read the old value of A or the new value of A? Justify your answer in each case.
   (b) Consider the four ACID properties of database transactions. In the above scenario, are these properties being guaranteed or not? Answer separately for each property, and justify.

11. [1.5 pts] Suppose that you have a data warehouse with sales by customer, product, and time. Also, suppose that you have an SQL query to calculate sales by customer country, product category, and year:

    select sum(sales.price), customer.country, product.category, time.year
    from sales **natural join** customer **natural join** product **natural join** time
    group by customer.country, product.category, time.year

    How would you change this query in order to do the following:
   (a) Drill-down to customer city.
   (b) Slice on year in order to show the sales for 2015 only.

Sistemas de Informação e Bases de Dados

2017-01-20

Answer the questions in the exam sheet.

Identify all sheets with your name and student number

You may answer the questions in English or Portuguese.

Duration: 90 minutes

Number: Name:

Part I: E-R modeling

Técnico Sub-Aqua Club (TSAC) is a hypothetical scuba diving club that needs your expertise to create its own database and, for that purpose, puts you in context using the following description:

Dive sites are identified by a name and have a geographic location. TSAC is specialized in two types of dive sites: wreck and cave. Any visited dive site belongs to one, and only one, of these types. A wreck is also characterized by a date, whereas a cave has a length and a height. A dive is done by a group of at least one diver who visits a dive site. A dive site may be visited in multiple dives. A dive has a start date/time, a time duration, and a unique URL of the dive plan. A diver is identified by his ID card number.

1. [2.0 pts] Draw an E-R diagram for this scenario. Include appropriate constraints in each association.

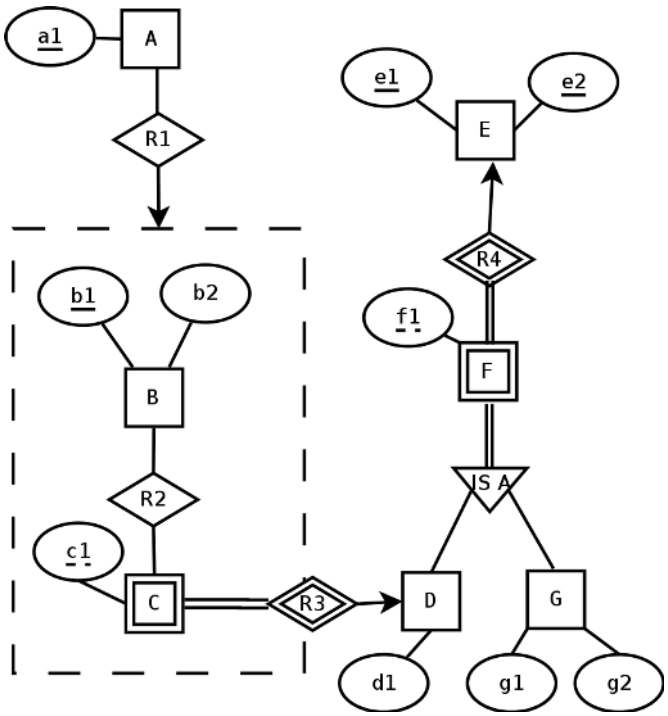
2. [2.0 pts] The following requirements have to be included in the database design:

   In a dive one diver is the coordinator. A dive is logged as a sequence of depth and temperature values captured at certain elapsed times after the dive start date/time.

   Each diver participates in a dive using one oxygen cylinder filled with compressed air/gas. A cylinder has a unique identification number and a capacity.

   Extend the previous E-R diagram to incorporate the above requirements. You only need to draw the additional elements.

3. [1.5 pts] Convert the following E-R diagram to the relational model.



Part II: SQL and PHP

We have a clinical database with info on patients, device used at a diabetes clinic and glucose level data acquisitions performed by patients using the available devices. Consider the schema of three database tables:

Patient(<u>patID</u>, name, birthyear)
Device(<u>devID</u>, manufacturer)
Reading(<u>devID</u>, <u>patID</u>, date, level)
devID: FK(Device)
patID: FK(Patient)

4. [3.0 pts] Write an SQL query for each of the following questions:
   a. How many readings have been acquired by devices of manufacturer “Medtronic” today? (i.e. on the day the query is performed)
   b. How many patients have never obtained a reading with level > 100?
   c. Which patient has obtained at least one reading with level > 70 from all devices of manufacturer "Medtronic"?

5. [1.5 pts] The SQL function *AVG_readings* returns the average glucose level on readings of the patient on a certain date:

   AVG_readings(patient, date)

   Write a query using the above function to return the names and year of birth of patients with the lowest average reading in 2017-01-01.

6. [1.5 pts] Write the code of a HTML form to register a new device (will correspond to inserting a new entry in the table Device above). The design of the form should restrict users to specifying manufacturers "Philips" or "Medtronic" only. The form should invoke PHP script "addDevice.php" when submitted. Also, the form must be designed in such way that, upon submission, the form data must not be passed in the URL of the server-side programme handling the form.

7. [1.5 pt] Write the PHP code of of addDevice.php to insert a new device into table Devices, when run on a Web server receiving the form of the previous question. Make sure your code prevents SQL injection.

Part III: Advanced topics

Suppose that you have a database with medical appointments. The following tables store each appointment between a patient and a doctor, on a certain date and time:

Appointment(<u>patient_id, doctor_id, date_time</u>)
FK(patient_id): Patient
FK(doctor_id): Doctor

Patient(<u>patient_id</u>, country)

Doctor(<u>doctor_id</u>, specialty)

Each appointment is supposed to last 20 minutes and there is a mandatory **condition**:

No appointment should be scheduled less than 20 minutes after the previous appointment with the same doctor.

8. [1.5 pts] Write an insert trigger to check the condition.
   Hint: You can use the function `TIMESTAMPDIFF(MINUTE, timestamp1, timestamp2)` to calculate the difference (*timestamp2 – timestamp1*) in minutes.

9. [1.0 pt] There will be situations when two or more appointments are being created at the same time, possibly for the same doctor. What is the lowest isolation level that should be used by concurrent transactions to ensure that the condition above is always guaranteed? Justify your answer.

10. [1.5 pts] Instead of using a trigger, we want to enforce the condition with a transaction implemented in PHP/PDO. Explain how this would be implemented. Indicate all the necessary PHP/PDO functions, and briefly describe their purpose.

11. [1.5 pts] Someone had the idea of creating an index on the doctor_id column. For each query below, would the query become faster with the index? Justify your answers.
    a. select patient_id from appointment where doctor_id in (513, 306);
    b. select doctor_id from appointment where month(date_time) = 8;
    c. select date_time from appointment where doctor_id >= 500;

12. [1.5 pts] The `appointment` table above can be seen as the fact table of a star schema with `patient` and `doctor` as dimension tables. Each patient has a country, and each doctor has a specialty.
    a. Write a query to show the number of appointments by country and specialty in 2016.
    b. Modify the query to also show the number of appointments for the rollup of country and specialty.

TÉCNICO LISBOA

Sistemas de Informação e Bases de Dados

20/1/2020

Evaluation Exam (duration: 2h30)

Student Name: _____ **Student Number:** _____

Group 1

The fitness company named "Conan e os Bárbaros" decided to use a relational database to support its day-to-day operations, storing all relevant data for its members/clients and trainers. You were hired to develop the database!

1a) Represent the following description for the information domain, using an Entity-Association (E-A) diagram:

> *One of the aims of the database is to record information about which members participated in which training sessions, taking place at different centers, and which trainers supervised which sessions.*
>
> *In more detail, the company operates various fitness centers in multiple cities. Every center is characterized by a unique name (e.g., "Muscle Maniacs", "Pump Factory", etc.), having also an address and one or more rooms, each with a unique number within the center (e.g., center 1, 2, etc.) and a maximum capacity.*
>
> *People can register for individual (i.e., independent) or group sessions in different centers. Each group session requires exactly one trainer, whereas individual sessions are done without a trainer.*
>
> *For each person of interest, including members and trainers, we want to store the first name, family name, and birth date. You can assume that the combination of first name, family name, and birth date is unique. For each trainer, the diploma is also recorded, besides the attributes that characterize all other individuals.*
>
> *A person can be a trainer in one session and participant in another (for individual or group sessions). The model should also include information about people (e.g., prospects) that have not participated in any sessions yet, or trainers (e.g., interns) that have not supervised any group sessions yet.*
>
> *For each session, the date and starting hour should be recorded. For group sessions, the type should also be stored (e.g., hatha yoga, aerobics, pumping iron, etc.). Different sessions can start at the same time on the same day, but in different centers or different rooms of a same center. At a given start hour of a day, at most one session can start in a given room of a center.* (2v)

1b) Consider the following statements as additions to the previous description for the information domain:

> *Sessions, either individual or in group, can involve equipment (e.g., treadmills, stationary bikes, weight machines and free weights, etc.), and each fitness center has a set of machines and other equipment that can be used. It is important to unambiguously known which fitness center has which equipment pieces, and you can consider that all equipment pieces have been used at least in one session.*
>
> *For each piece of equipment owned by the company, the database should store a unique identifier, a name, a type (e.g., stationary bikes, weight machines, etc.), the brand, and the date of acquisition. Each session can involve multiple pieces of equipment from the fitness center.*
>
> *Some of the trainers working for the company are specialists (i.e., certified) in the use of particular types of equipment. This information should be stored in the database, together with a description for the certification. Trainers can be specialists in multiple types of equipment.*

Represent the corresponding E-A model. It is not necessary to repeat the elements that remain equal to those from the previous question. Indicate also, as textual notes, all constraints not captured in the model. *(1.5v)*

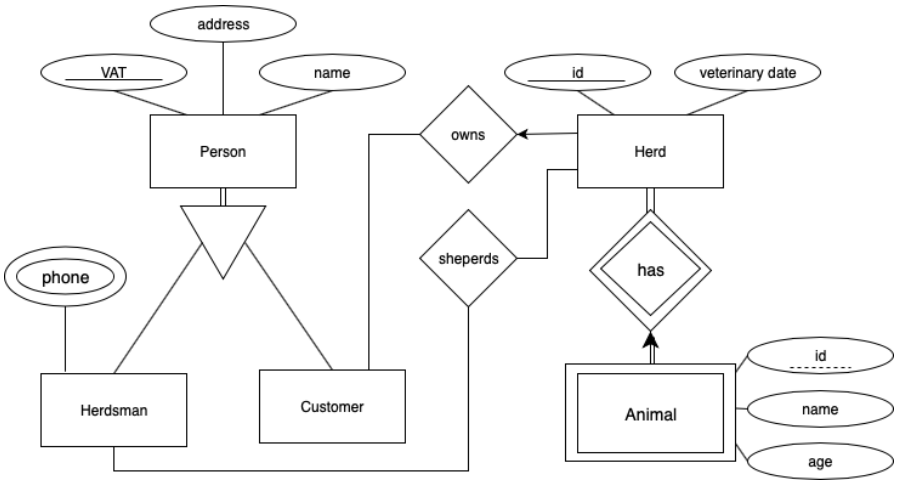
1c) Consider the following statements as additions to the previous description for the information domain:

> *The database should store additional information for a particular type of sessions, corresponding to physiotherapy sessions. In these cases, contrarily to normal individual or group sessions, only a single individual registers to attend, and a trainer is also mandatorily involved.*
>
> *For physiotherapy sessions, we also store the name of the doctor that prescribed the session, together with the member's health insurance code.*

Represent the corresponding E-A model. It is not necessary to repeat the elements that remain equal to those from the previous question. Indicate also, as textual notes, all constraints not captured in the model. *(0.5v)*

Group 2

Convert the following Entity-Association (E-A) model to the relational model. Do not forget to **describe/present integrity constraints for aspects that cannot be captured in the relational model**.



Use the following notation to denote primary keys and foreign keys. *(4v)*

$relation\_A(\underline{attribute\_1}, attribute\_2, ...)$
$attribute\_2 : FK(relation\_B)$

Group 3

Consider that the "No Pain, No Gain" fitness company relies on a relational database to store information about its activities, using the following relations as part of the database schema.

```
gym ( phone , name , district , city , city_borough )

instructor ( name , specialization , gender , hourly_salary , city , address )

works ( name , phone , date , hours )
  name: FK(instructor)
  phone: FK(gym)
```

Present a single SQL statement for each of the following information needs:

3a) What are the names of the male instructors with a specialization containing the word "yoga" and that work on gyms in the city of Lisbon? *(2v)*

3b) Which female instructors living in the city of Lisbon have, in total, already earned more than 5000 Euros from the company, from all the hours that they have worked in the different gyms? *(1.5v)*

3c) What is(are) the gym(s), in the city of Lisbon, associated with the largest number of "zumba" instructors? *(0.5v)*

Group 4

Considering the database from the exercises of Group 3, present SQL statements for each of the following questions.

4a) In the context of a single transaction, we aim to perform the following set of operations:

1. Increase the salary of instructors that worked at gyms in different cities from where they live, by 10%.
2. For the remaining instructors, increase their salary also by 10% if they are female.
3. Remove instructors not living in the same city of a gym where they have worked more than 1000 hours.

Show the SQL statements that are involved. *(2v)*

4b) Present a single SQL query that, using the OLAP grouping options introduced in the classes, can count the total number of gyms, and the number of gyms per location. Consider the administrative hierarchy of locations (i.e., borough ← city ← district) for presenting subtotals and totals for the counts. *(1.5v)*

4c) Create a trigger to prevent the insertion of tuples, in the "works" relation, for cases in which the same instructor has, in total, been paid less for other dates in the same gym. *(0.5v)*

Group 5

Consider the database used in the exercises from Group 3. Answer each of the following questions:

5a) In the relation named "instructor", consider that there is a functional dependency "specialization $\mapsto$ hourly_salary". Assuming that all attribute values are atomic, state in which normal form is the relation named "instructor" and, if needed, present a lossless-join and dependency preserving BCNF decomposition.

State also in which normal form would the "instructor" relation be if the attribute named "specialization" was part of the primary key for the relation. If the normal form is different, explain why. *(2v)*

5b) Consider SQL queries for retrieving (a) the instructors with an hourly salary below 50 Euros per hour, and (b) the instructors specializing in "zumba". For each query, describe an index that could improve its performance, and assume than several instructors (perhaps many) earn less than 50 Euros per hour. Display the SQL commands associated with the index creation, and justify your response, stating if the indexes should be based on B+trees or hashing, and if they should be primary or secondary. *(1.5v)*

5c) Explain what "SQL injection flaws" are and how they can be prevented in PHP Web applications. *(0.5v)*

Use this sheet for you drafts. You can separate it from the rest of the test.

–draft–

TÉCNICO LISBOA

Sistemas de Informação e Bases de Dados

18/1/2019

Evaluation Exam (duration: 2h30)

Student Name: _____ **Student Number:** _____

Group 1

Europe needs entrepreneurs! As such, we will be creating a completely innovative pastoral consulting company in Portugal, named "Vá Pastar S.A."

1a) Represent the following description for the information domain, using an Entity-Association (E-A) diagram:

> *The company must keep information about customers (i.e., the owners of the animals), herdsmen (i.e., the shepherds attending to animals), herds and animals.*
>
> *For customers and herdsmen, we should know the name, VAT, and address. Additionally, for herdsmen, we need to store information on phone numbers (possibly more than one). For animals, we must know the name, identification number, and age. For herds, we should know the identification number and the date of the last veterinary inspection.*
>
> *A customer may own several herds and each herd may belong to several customers. A herdsman can shepherd several herds and each herd may be shepherded by several herdsmen. Each animal belongs to a single herd and a herd can have several animals (at least one).*

(2v)

1b) Consider the following statements as additions to the previous description for the information domain:

> *There are only 3 types of animals: cows, sheep and goats. For cows, it is necessary to store the average of milk that is produced. For sheep, it is necessary to keep the different dates on which shears were made. For goats, it is necessary to store the gender (i.e., male or female).*
>
> *The company also keeps information on existing pastures. For each pasture it is necessary to know its GPS coordinates of latitude and longitude. When a herd is shepherded, in addition to knowing who the herdsman was, it is necessary to know which pasture is used and on what date(s).*

Represent the corresponding E-A model. It is not necessary to repeat the elements that remain equal to those from the previous question. *(1.5v)*

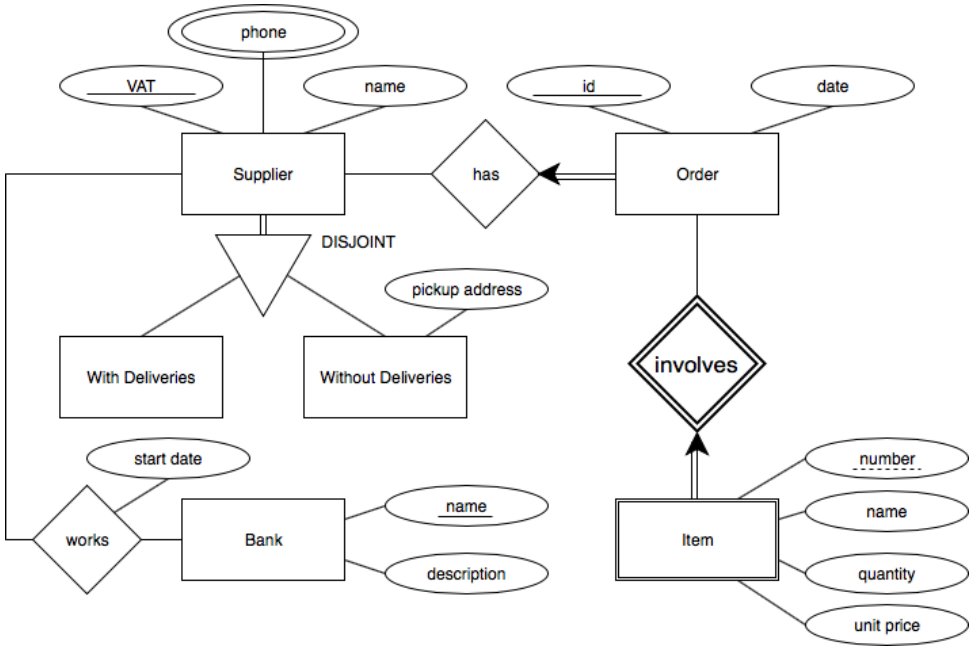
1c) Consider the following statements as additions to the previous description for the information domain:

> *Some of the herdsmen of the company also assume the role of inspectors. The inspectors sporadically verify the performance of other herdsmen.*
> *Moreover, pastures are divided into sectors. These are identified by a letter (e.g., sector A of a given pasture p_1, sector B of pasture p_1, sector A of pasture p_2, etc.) and it is necessary to save their area.*

Represent the corresponding E-A model. It is not necessary to repeat the elements that remain equal to those from the previous question. *(0.5v)*

Group 2

Convert the following Entity-Association (E-A) model to the relational model.



Use the following notation to denote primary keys and foreign keys. *(4v)*

$relation\_A(\underline{attribute\_1}, attribute\_2, ...)$
$attribute\_2 : FK(relation\_B)$

Group 3

The home repair company named "Aluga-se marido" registers information about its employees and the works (i.e., projects) it carries out. Consider that the company uses a relational database to store information about its activities, using the following relations as part of the database schema:

```
technician ( VAT , name , type , city , priceHour )

project ( id , address , city )

works ( VAT , id , hours )
  VAT: FK(technician)
  id:  FK(project)
```

Present a single SQL statement for each of the following information needs:

3a) What are the names of the technicians with type "painter" that work on projects in the city of Lisbon? *(2v)*

3b) Which technicians have, in total, already worked more than 500 hours for the company? *(1.5v)*

3c) What is(are) the project(s) in the city of Lisbon associated with a higher total cost? *(0.5v)*

Group 4

Consider the database used in the exercises from Group 3. Present SQL statements for each of the following questions:

4a) In the context of a single transaction, we aim to perform the following set of operations:

1. Raise the hourly price associated to technicians of the type "plumber" by 10%.
2. Remove from the database all technicians who have not yet worked on company projects, and whose price per hour exceeds 100 Euros (i..e, remove technicians matching the two conditions).

Show the SQL statements that are involved. *(2v)*

4b) Insert in the database a new technician of the type "bricklayer", with a name and VAT defined by you. The price per hour should be equal to the average price of the "bricklayer" technicians in the database. *(1.5v)*

4c) What are the names and VATs of the technicians who worked on all the projects from the company? *(0.5v)*

Group 5

Consider the database used in the exercises from Group 3. Consider also that the different tables have no indices, and that several thousand records are stored on each table. Answer each of the following questions:

5a) In the relation named "technician", consider that there is a functional dependency "type $\mapsto$ priceHour". Assuming that all attribute values are atomic, state in which normal form is the relation named "technician" and, if needed, present a lossless-join and dependency preserving BCNF decomposition. *(2v)*

5b) If you were to create a primary index on the "project" table, what column(s) would you create it on? Display the SQL command associated with the index creation, and justify your response, giving a SQL query that would clearly benefit from the index. *(1.5v)*

5c) State whether the index suggested for the previous question could be based on hashing (i.e., indicate if a hash index could also be useful for the SQL query given for the previous question). Justify your answer. *(0.5v)*

Use this page for you drafts.

Use this sheet for you drafts. You can separate it from the rest of the test.

–draft–

Sistemas de Informação e Bases de Dados

2017-01-19

Answer the questions in the exam sheet.

Identify all sheets with your name and student number

You may answer the questions in English or Portuguese.

Duration: 90 minutes

Number: Name:

Part I: E-R modeling

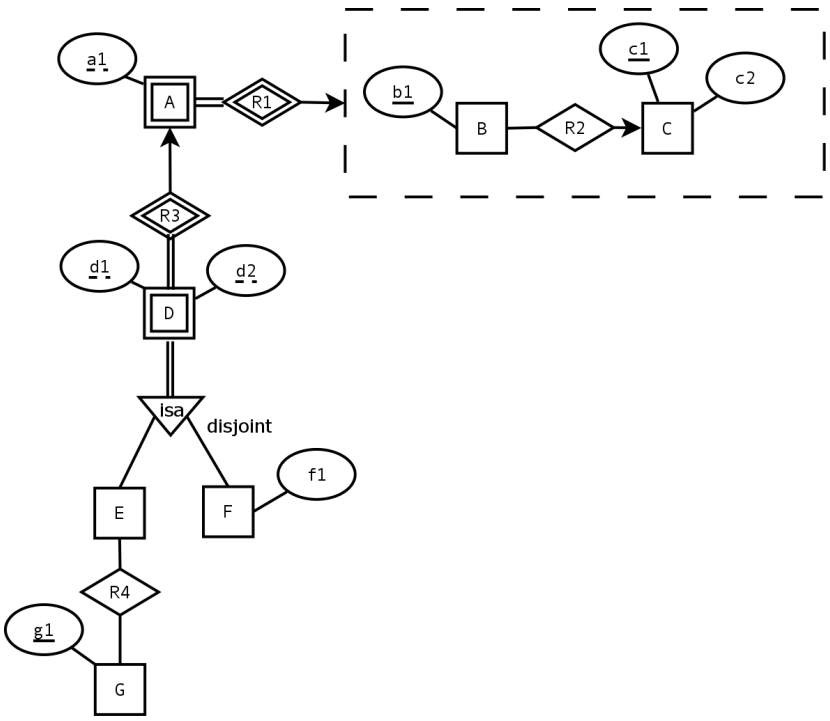
A hospital aims to reorganize itself. Your role is to create a conceptual model to meet the following description of the hospital's clinical operation. The primordial operational concept corresponds to clinical event. A clinical event combines an order with one performed service, where each of these concepts has its own specific identification code. A clinical event must follow a diagnostic procedure. Diagnostic procedures are named and universally identified through a specific code. Also, each diagnostic procedure has a specific set of possible test result types, which themselves are uniquely identified by a type code. Each clinical event comprehends a sequence of clinical event components, which are incrementally numbered within the context of the clinical event. Finally, each clinical event component returns a test result type that must match one of the type codes for the diagnostic procedure being followed.

1. [2.0 pts] Draw an E-R diagram for this scenario. Include appropriate constraints in each association and any additional notes if necessary.

2. [2.0 pts] Consider two extensions to the ER model.
    a. A clinical event involves a patient and a professional, both identified by a Citizen ID number. The professional and the patient must not be the same person.
    b. The performed service should not be identified by a specific code any more. In fact, a service is performed at a given date/time, in a location (with specific ID code), and using an equipment (with specific ID code).

Modify the previous E-R diagram to incorporate the above extensions. You only need to draw the changed and new elements.

3. [1.5 pts] Convert the following E-R diagram to the relational model.



Number: Name:

Part II: SQL and PHP

We have a chemical database with info on atoms, molecules, chemical elements and isotopes.
Consider the schema, where columns *protons* and *neutrons* are *integers*, and all other columns are *strings*:

Isotope(<u>protons</u>, <u>neutrons</u>)
protons: FK(Element)

Element(<u>protons</u>, symbol, name)

Molecule(<u>formula</u>, name)

Atom(<u>AtomID</u>, formula, protons, neutrons)
formula: FK(Molecule)
protons, neutrons: FK(Isotope)

To insert a molecule of "Water", "H2O", in the schema an SQL script as this one would do it:

```
Insert into Element values(1, 'H', "Hydrogen"); Insert into Element values(8, 'O', 'Oxygen');
Insert into Isotope values(1,1); Insert into Isotope values(8,8);
Insert into Molecule values('H2O', 'Water');
Insert into Atom('A1', 'H20', 1, 1); Insert into Atom('A2', 'H20', 1, 1); Insert into Atom('A3', 'H20', 8, 8);
```

4. [3.0 pts] Write an SQL query for each of the following questions:
   a. How many molecules in the database have at least an atom of "Oxygen"?
   b. What are the names of the molecules in the database that contain at least an atom of "Carbon" and no "Oxygen"?
   c. Which elements have all their isotopes in molecules of the database?

5. [1.5 pts] Write an SQL Function Mass(f) that returns the atomic mass of a molecule given its formula (hint: the atomic mass of a molecule is the count of neutrons and protons of all the atoms in the molecule).

6. [1.5 pts] Write a query that invokes the above function to return the name of the molecule having “acid” in its name that has the highest atomic mass.

7. [2.0 pts] HTML and PHP trivia. Write the code to register a new chemical element in the database (i.e., inserting a new entry in the table Element).
    a. Write the HTML code to generate a web browser form to register a new chemical element. Make sure that the data collected on the form will not be transferred to the Web server in the clear.
    b. Write the PHP code to process the above form. Validate that “protons” must be a number between 1 and 54. Make sure your code prevents SQL injection.

Part III: Advanced topics

We are referring to the same schema as in Part II.

Isotope(<u>protons</u>, <u>neutrons</u>)
protons: FK(Element)
Element(<u>protons</u>, symbol, name)
Molecule(<u>formula</u>, name)

Atom(<u>AtomID</u>, formula, protons, neutrons)
formula: FK(Molecule)
protons, neutrons: FK(isotope)

8. [2.0 pts] General questions on transactions.
   a. If a transaction operates at the "repeatable reads" isolation level can it happen that we read uncommitted data from a table, when other transactions do not operate at the "serializable" isolation level? Why?
   b. Explain what kind of anomaly can be observed when scanning a table at the "repeatable reads" isolation level.

9. [1.5 pts] Write a PHP script to insert a molecule of "Water" (and its Atoms), "H2O", composed of isotopes with protons, neutrons=1,1 and 8,8. Add the molecule data only if it is not already in the database and make sure that the script either adds all rows or none.

10. [1.5 pts] When the Atom table reached 10 million rows, the DB Administrator created additional data structures on the database with these 2 commands:

    create index protons_ix on Atom(protons) using BTREE
    create index formula on Atom(formula) using HASH

    For each SQL instruction below, would the query become faster or slower with the created indexes? Justify your answers.

    a. insert into Atom('B1', 'NaCl', 11, 12)
    b. select count(*) from Atom where formula='NaCl'
    c. elect distinct formula from Atom where neutrons=12

11. [1.5 pts] Let's look at the schema of the Chem Database as a data warehouse organized as a star schema, in which the facts are the atoms and the dimensions are molecule, isotope/element.

    a. Which columns in the fact table represent each of dimensions?
    b. Write a query to show the number of atoms of element Carbon by molecule and isotope.
    c. Write a query to show the total number of atoms of every isotope of Carbon and the combined total of Carbon atoms in the database.

    Use the MySQL extension to SQL GROUP BY for warehousing when helpful.

TÉCNICO LISBOA

Sistemas de Informação e Bases de Dados

18/1/2019

Evaluation Exam (duration: 2h30)

Student Name: _____ **Student Number:** _____

Group 1

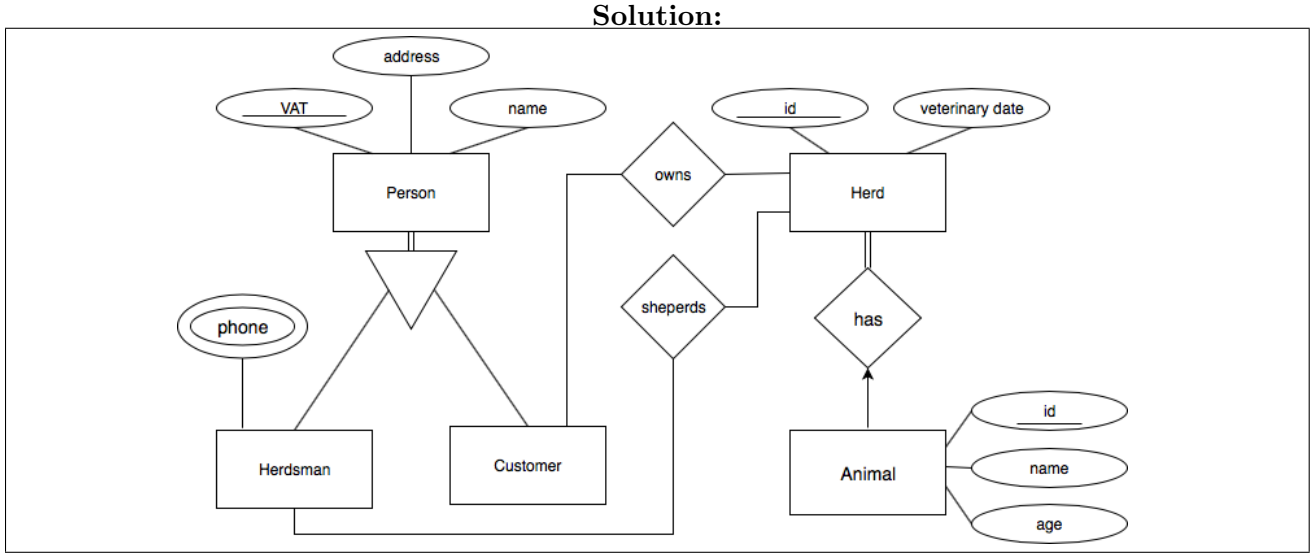
Europe needs entrepreneurs! As such, we will be creating a completely innovative pastoral consulting company in Portugal, named "Vá Pastar S.A."

1a) Represent the following description for the information domain, using an Entity-Association (E-A) diagram:

> *The company must keep information about customers (i.e., the owners of the animals), herdsmen (i.e., the shepherds attending to animals), herds and animals.*
>
> *For customers and herdsmen, we should know the name, VAT, and address. Additionally, for herdsmen, we need to store information on phone numbers (possibly more than one). For animals, we must know the name, identification number, and age. For herds, we should know the identification number and the date of the last veterinary inspection.*
>
> *A customer may own several herds and each herd may belong to several customers. A herdsman can shepherd several herds and each herd may be shepherded by several herdsmen. Each animal belongs to a single herd and a herd can have several animals (at least one).*

(2v)

Solution:

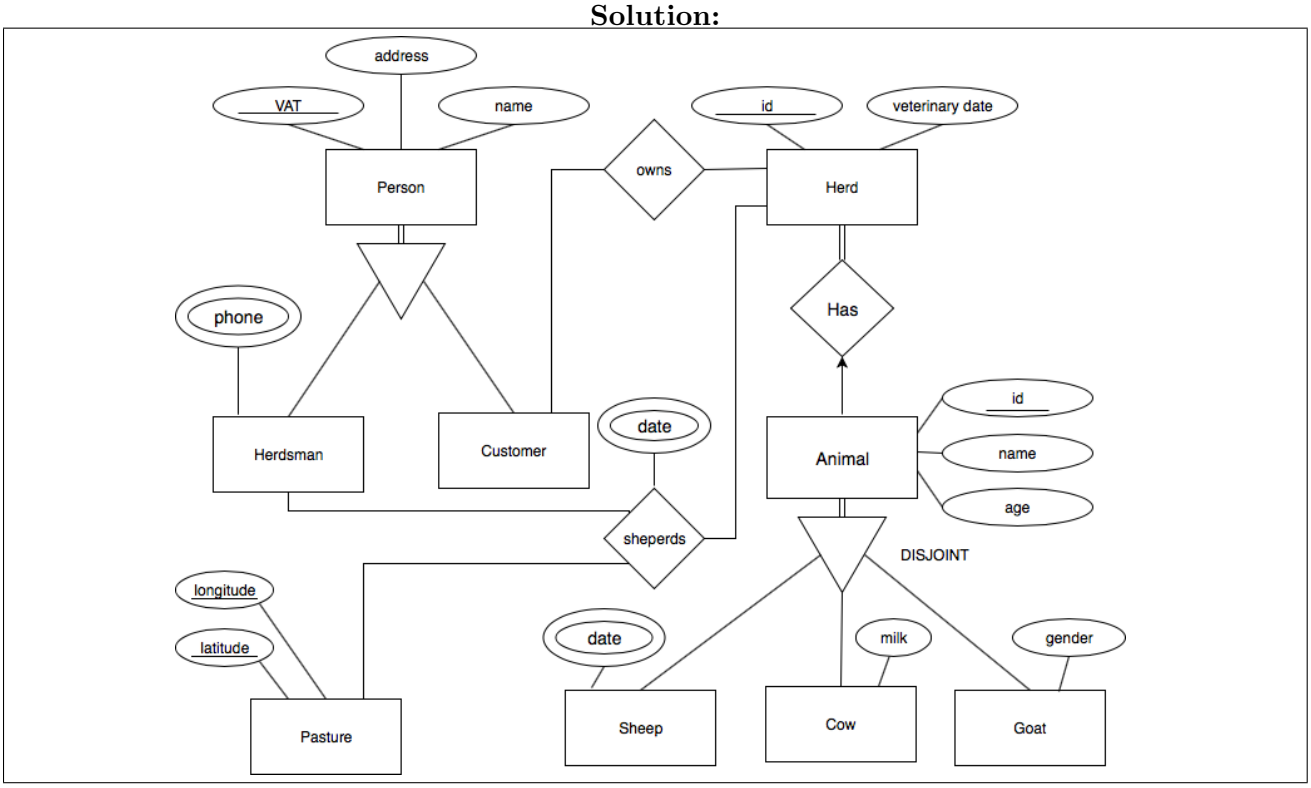


1b) Consider the following statements as additions to the previous description for the information domain:

> *There are only 3 types of animals: cows, sheep and goats. For cows, it is necessary to store the average of milk that is produced. For sheep, it is necessary to keep the different dates on which shears were made. For goats, it is necessary to store the gender (i.e., male or female).*
>
> *The company also keeps information on existing pastures. For each pasture it is necessary to know its GPS coordinates of latitude and longitude. When a herd is shepherded, in addition to knowing who the herdsman was, it is necessary to know which pasture is used and on what date(s).*

Represent the corresponding E-A model. It is not necessary to repeat the elements that remain equal to those from the previous question. *(1.5v)*

Solution:

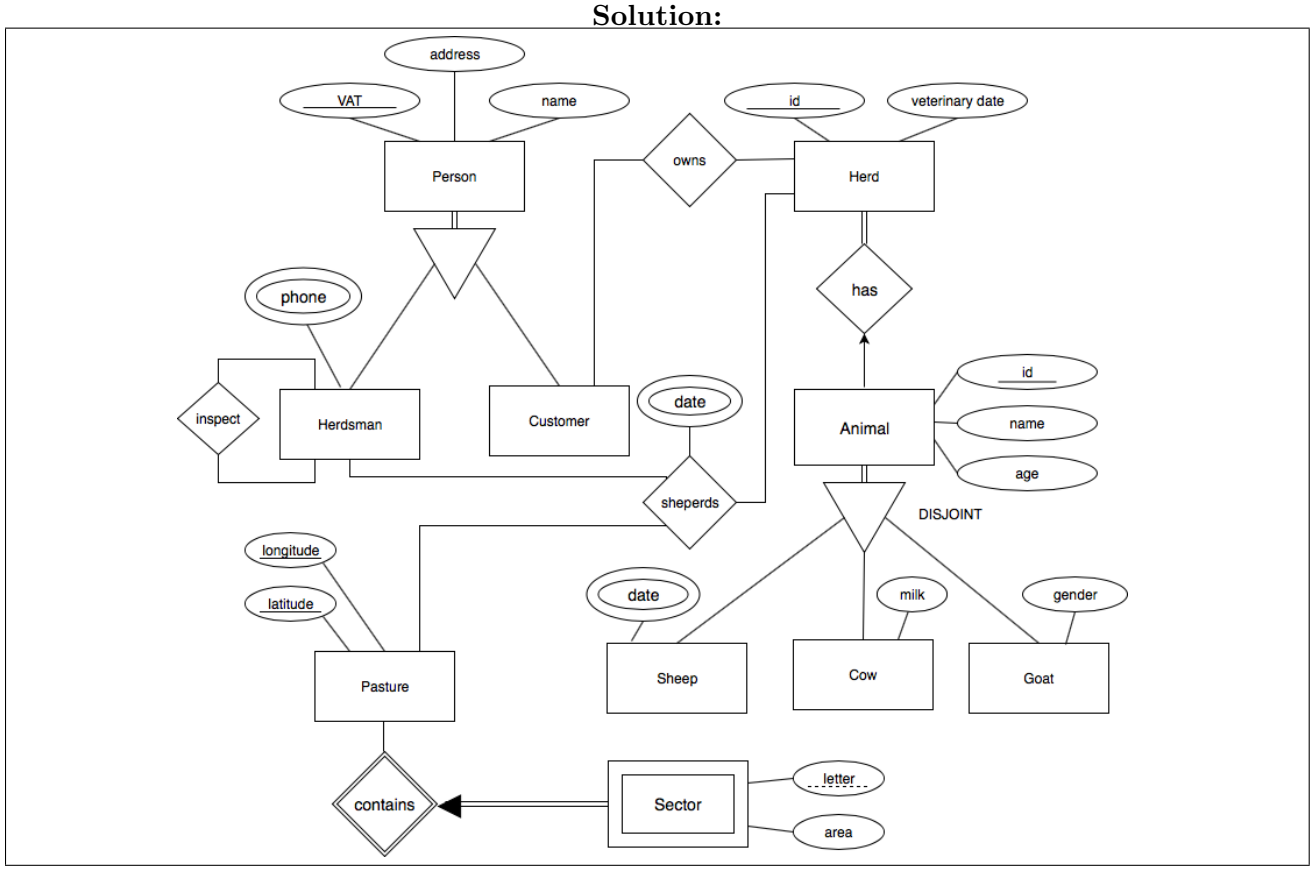


1c) Consider the following statements as additions to the previous description for the information domain:

> *Some of the herdsmen of the company also assume the role of inspectors. The inspectors sporadically verify the performance of other herdsmen.*
> *Moreover, pastures are divided into sectors. These are identified by a letter (e.g., sector A of a given pasture p_1, sector B of pasture p_1, sector A of pasture p_2, etc.) and it is necessary to save their area.*

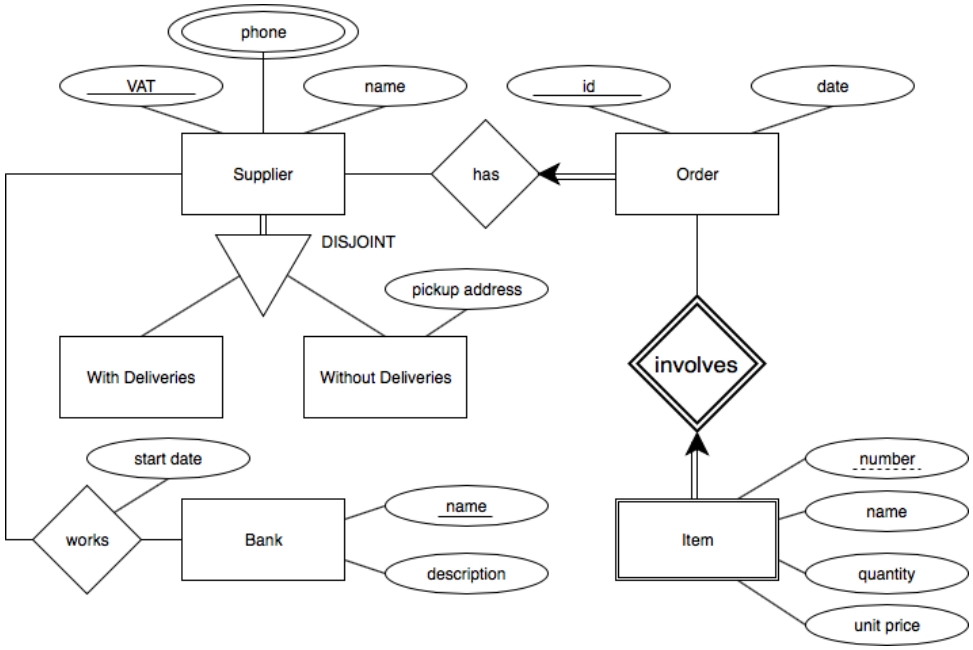
Represent the corresponding E-A model. It is not necessary to repeat the elements that remain equal to those from the previous question. *(0.5v)*

Solution:



Group 2

Convert the following Entity-Association (E-A) model to the relational model.



Use the following notation to denote primary keys and foreign keys. *(4v)*

$$relation\_A(\underline{attribute\_1}, attribute\_2, ...)$$
$$attribute\_2 : FK(relation\_B)$$

Solution:

$Supplier(\underline{VAT}, name)$
IC: All suppliers are either "with deliveries"or "without deliveries", and they cannot be both options simultaneously.
IC: All suppliers need to have a phone number.

$$Phone(\underline{VAT, phone})$$
$$VAT : FK(Supplier)$$

$$WithDeliveries(\underline{VAT})$$
$$VAT : FK(Supplier)$$

$$WithoutDevliveries(\underline{VAT}, pickup\_address)$$
$$VAT : FK(Supplier)$$

$$Bank(\underline{name}, description)$$

$$Order(\underline{id}, date, VAT)$$
$$VAT : FK(Supplier)$$

$$Item(\underline{id, number}, name, quantity, unit\_price)$$
$$id : FK(Order)$$

$$Works(\underline{name, VAT}, start\_date)$$
$$VAT : FK(Supplier)$$
$$name : FK(Bank)$$

Group 3

The home repair company named "Aluga-se marido" registers information about its employees and the works (i.e., projects) it carries out. Consider that the company uses a relational database to store information about its activities, using the following relations as part of the database schema.

```
technician ( VAT , name , type , city , priceHour )

project ( id , address , city )

works ( VAT , id , hours )
  VAT: FK(technician)
  id:  FK(project)
```

Present a single SQL statement for each of the following information needs:

3a) What are the names of the technicians with type "painter" that work on projects in the city of Lisbon? *(2v)*

Solution:

```
SELECT name
FROM technician, project, works
WHERE type="painter"
 AND project.city="Lisbon"
 AND technician.VAT = works.VAT
 AND project.id = works.id
```

3b) Which technicians have, in total, already worked more than 500 hours for the company? *(1.5v)*

Solution:

```
SELECT VAT, name
FROM technician, works
WHERE technician.VAT = works.VAT
GROUP BY VAT, name
HAVING SUM(hours) > 500
```

3c) What is(are) the project(s) in the city of Lisbon associated with a higher total cost? *(0.5v)*

Solution:

```
SELECT id
FROM project, works, technician
WHERE project.city="Lisbon"
AND project.id=works.id
AND technician.VAT=works.VAT
GROUP BY id
HAVING SUM(hours * priceHour) >= ALL (
  SELECT SUM(hour * priceHour)
  FROM project, works, technician
  WHERE project.city="Lisbon"
  AND project.id=works.id
  AND technician.VAT=works.VAT
  GROUP BY id
)
```

Group 4

Consider the database used in the exercises from Group 3. Present SQL statements for each of the following questions:

4a) In the context of a single transaction, we aim to perform the following set of operations:

1. Raise the hourly price associated to technicians of the type "plumber" by 10%.
2. Remove from the database all technicians who have not yet worked on company projects, and whose price per hour exceeds 100 Euros (i..e, remove technicians matching the two conditions).

Show the SQL statements that are involved. *(2v)*

Solution:

```
BEGIN TRANSACTION;

UPDATE technician SET priceHour = priceHour * 1.1
WHERE type="plumber";

DELETE FROM technician
WHERE priceHour > 100
AND VAT NOT IN ( SELECT VAT FROM works );

COMMIT;
```

4b) Insert in the database a new technician of the type "bricklayer", with a name and VAT defined by you. The price per hour should be equal to the average price of the "bricklayer" technicians in the database. *(1.5v)*

Solution:

```
INSERT INTO technician (
  SELECT 12345, "John", "bricklayer", AVG(priceHour)
  FROM technician
  WHERE type="bricklayer"
)
```

4c) What are the names and VATs of the technicians who worked on all the projects from the company? *(0.5v)*

Solution:

```
SELECT VAT, name
FROM technician
WHERE NOT EXISTS (
  SELECT id
  FROM PROJECT
  WHERE id NOT IN (
    SELECT id
    FROM works
    WHERE works.VAT=technician.VAT
  )
);
```

Group 5

Consider the database used in the exercises from Group 3. Consider also that the different tables have no indices, and that several thousand records are stored on each table. Answer each of the following questions:

5a) In the relation named "technician", consider that there is a functional dependency "type $\mapsto$ priceHour". Assuming that all attribute values are atomic, state in which normal form is the relation named "technician" and, if needed, present a lossless-join and dependency preserving BCNF decomposition. *(2v)*

Solution:

There's a single candidate key and a single key attribute, namely VAT.

The relation is in 1NF because all attribute values are atomic.
The relation is in 2NF because all non-key attributes depend totally on the candidate keys.
The relation is not in 3FN because there are dependencies between non-key attributes.

A lossless BCNF decomposition would be:

```
technician ( VAT , name , type , city )
technician_types ( type , priceHour )
```

The decomposition is lossless-join because the common attribute in the new relations is a candidate key in one of the new relations.

5b) If you were to create a primary index on the "project" table, what column(s) would you create it on? Display the SQL command associated with the index creation, and justify your response, giving a SQL query that would clearly benefit from the index. *(1.5v)*

Solution:

An index could be created on the attribute project.id, which corresponds to the primary key of the table.

```
CREATE INDEX idx1 ON project (id);
```

This index supports the optimization of joins with the project table, usually involving this attribute, or selection queries using conditions on project.id, such as:

```
SELECT * FROM obra WHERE id > 100;
```

5c) State whether the index suggested for the previous question could be based on hashing (i.e., indicate if a hash index could also be useful for the SQL query given for the previous question). Justify your answer. *(0.5v)*

Solution:

It would not be useful for the query of the previous exercise, since this query involves a search for a range of values. Hash indexes do not support searches for ranges of values.

Use this page for you drafts.

Use this sheet for you drafts. You can separate it from the rest of the test.

–draft–

Sistemas de Informação e Bases de Dados

Exame de 1ª Época (Época Normal)

This exam is written in English, but you may answer in Portuguese.

Introduction – Consider the following scenario:

A health club offers several activities (e.g. yoga, cycling, aerobics, etc.), which are organized into classes. For example, there is an aerobics class at 8:00, a yoga class at 9:00, etc. It is assumed that classes start every hour and have a 1-hour duration. There may be multiple classes for the same activity during the day (e.g. one yoga class in the morning, and another in the afternoon). Each class is given by a particular instructor. An instructor is just an employee, who is identified by name and has a job description and a birthday. On the other hand, the club members are identified by number, but also have a name and birthday. Club members can participate in any class they want, and it is necessary to keep track of which members participate in which class.

An IT company was hired to create a database for this health club. Here is their solution:

member(<u>number</u>, name, birthday)

class(<u>time, activity</u>, name)
name : FK(employee)

participant(<u>number, time, activity</u>)
number : FK(member)
time, activity : FK(class)

employee(<u>name</u>, description, birthday)

Part I: E-R **[Answer in a separate sheet of paper]**

1. From the solution proposed by the IT company, draw an E-R (Entity-Relationship) model that captures their conceptual view of this scenario.

2. Both employees and members of this health club have a name and a birthday. Is it possible to create a more general entity called *Person* and then use a specialization ("is-a") relationship for *member* and *employee*? Explain your answer.

3. Sometimes, a member can bring a friend (who is not a club member) to a class. For marketing purposes, the health club stores the friend's name, phone and email, and also wants to know which friends each member brought to each class. Draw an E-R diagram for this part of the scenario.

Part II: SQL **[Answer in a separate sheet of paper]**

4. Write an SQL query to determine which club members have the same birthday as some employee. Show the members and the employee in the result.

5. Write an SQL query to determine who is the oldest member in the health club.

6. Write an SQL query to determine which class has the largest number of participants.

7. Someone has noticed that the members who do cycling and the members who do aerobics are mutually exclusive (i.e. no single member does both cycling and aerobics). Write a SQL query to determine whether this is true or not.

8. Write a SQL query to determine which member participates in all 'yoga' classes.

9. Each employee has a job description (e.g. 'manager', 'trainer', 'receptionist', etc.). However, the instructor of a class must always be a 'trainer'. Write a trigger to make sure that this is always the case.

Part III: PHP [Answer in a separate sheet of paper]

Suppose that you have to create an HTML form and a PHP script to register (in the database) that a certain member participates in a certain class.

10. How would the HTML form look like? Draw what the user would see in the Web browser when opening the page.

11. How does the PHP script receive the data from the HTML form? Explain how this mechanism works (in general).

12. What is the SQL query and the PHP functions that you would use to interact with the database in the PHP script? Explain why you need each PHP function in particular.

Part IV: Advanced Topics [Answer in a separate sheet of paper]

13. In case an instructor does not show up for a class, the health club needs to find a suitable replacement as quickly as possible. Basically, we need to find all the instructors who give a certain activity. How would you use indexes to optimize such query? Explain your answer.

14. What is an execution plan? What can you see in an execution plan? Is it possible to have different execution plans for the same query? How? Explain with your own words.

15. Suppose there is a limit to the number of participants in each class (for example, 30). How could you use transactions to ensure that such limit is never exceeded? Explain what the transaction would do. Also, give an example of how the limit could be exceeded without the use of transactions.

The health club has just opened a swimming pool and there are many members interested in swimming classes (about 100 members). The manager gets 3 employees on 3 different computers to use your HTML form (see exercise 10.) to store the list of participants (one by one) into the database. However, while the employees are doing this, the system crashes.

16. What will the database system do to recover from this situation? (What does a database system do, in general, to recover from a crash?)

17. When the system is up again, what should the employees do? Can they resume their work, or do they have to do everything again from the beginning? (Imagine you are the manager; what would you tell them to do?)

Sistemas de Informação e Bases de Dados

2016-01-30

Exam

You may answer the questions in English or Portuguese. Duration: 90 minutes

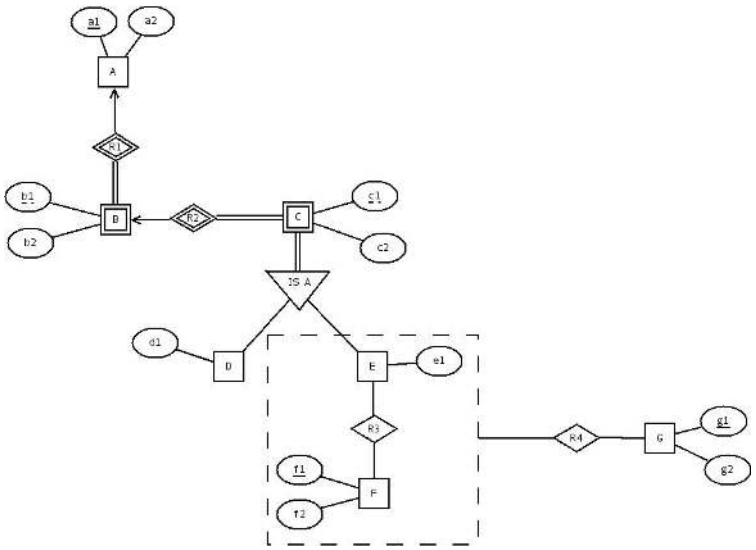
Part I: E-R modeling

You have been hired to design a database for an automobile repair shop. For this purpose, you are given the following description:

Vrum Limited is a multi-brand car servicing station. For every service, the database should store information about the car, namely the registration identifier, model, type of fuel used, and mileage recorded. In addition, every car has at least one driver, but a driver may have several cars. Each driver has a national ID, an address and a phone number. An appointment has to be scheduled for each car that needs to be serviced, and each appointment has a day and a time. It is important to note that while an appointment is always related to one car, a car may have more the one appointment scheduled with the servicing station. After completing the service of an appointment, a repair report should describe the work done; specifically, a report should have a list of repairs, where each repair specifies the part supplied, the cost of the part, and labor charges.

Answer the following questions in a separate sheet of paper

1. [2.5 pts] Draw an E-R diagram for this scenario. Include appropriate constraints in each association.
2. [2.0 pts] The following requirements have to be included in the database design:
   All the mechanics have an employment contract with the repair shop. The mechanics have a national id, an address and a phone number, while each contract is identified by a number and should include contract terms. Not every mechanic has a full-time contract, thus some of the mechanics have part-time contacts. Mechanics with a full-time contract have a list of benefits (e.g., annual leave, sick leave, and health insurance), while part-time contracts do not come with these benefits. A part-time contract must also specify the number of hours of work per week. The repair shop also keeps track of the mechanics that are allocated to each repair.
   Draw an update the previous E-R diagram incorporating the above requirements.
3. [1.5 pts] Convert the following E-R diagram to the relational model.



Part II: SQL and PHP

Consider a database for managing energy consumption in a large organization. There are energy meters scattered across several campi that record the count of energy units consumed (Kwh) at the end of every hour. There are 2 energy rates, A and B, for two periods of consecutive hours.

Meter

MeterID	Building	Department	Campus
001	NorthTower	EE	Alameda
002	NorthTower	EE	Alameda
003	SouthTower	BioEng	Alameda
101	ISTTagus	EE	Taguspark
102	InformaticaII	CSE	Alameda
201	Area51	EE	Nevada

Rate

RateClass	UnitPrice	StartHour	EndHour
A	0.20	08	22
B	0.10	23	07

Reading

MeterID	Month	Day	Hour	Units
001	12	25	00	1023
002	12	25	00	255
001	12	25	01	1024

Answer the following questions in a separate sheet of paper

4. [4.0 pts] Write an SQL query for each of the following questions:
   a) How many readings above 1000 units have been collected from meter 001 in Month 12?
   b) Which building had more readings in Month 12 at the highest UnitPrice? [ORDER BY clause not allowed]
   c) Which departments have buildings in all campi? [HAVING clause not allowed]

5. [2.0 pts] Write an SQL trigger for table Rate to guarantee that the UnitPrice of Rate B can never more than double the UnitPrice of Rate A.

6. [1.5 pts] Write the HTML the code to render the form of the figure and have a web client invoke the URL when "Go!" is clicked with the data also seen on the figure
(the "frame" enclosing the fields is optional).

http://h.pt/*addMeter.php?B="NorthTower"?D="EE"?C="Alameda*"

7. [1.5 pts] Write the PHP code of script *addMeter.php* to register a new meter in the Meter table of the database, given its location (i.e., attibutes Building, Department and Campus). The new MeterID should be the highest MeterID number in the database + 1. The location data is passed in URLs with the syntax of the URL of the previous question.

Part III: Advanced topics

Answer the following questions in a separate sheet of paper

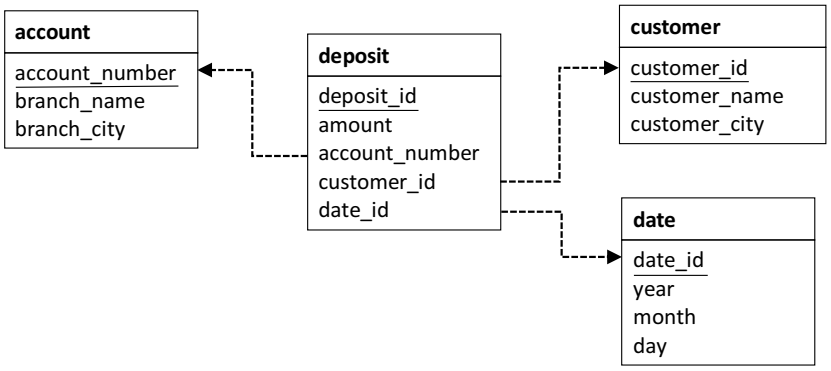
8. [2.0 pts] Consider the following Web page:

```
<form action="deposit.php">
<input type="hidden" name="customer_name" value="<?=$_SESSION['customer_name']?>"/>
<input type="hidden" name="account_number" value="<?=$_REQUEST['account_number']?>"/>
<input type="text" name="amount"/>
<input type="submit"/>
</form>
```

   (a) When the form is submitted, the script deposit.php needs to access the customer name, the account number, and the amount inputs. Which PHP array should the script use to access each of these inputs? Justify your answer.
   (b) Suppose that, in deposit.php, you want to add the account number and the amount to the session data. Write the code to do that.

9. [1.5 pts] Suppose that there are two transactions (T1 and T2) running in parallel in READ COMMITTED mode. At a certain point, T2 gets blocked (i.e. it "freezes") until T1 commits.
   (a) Give a sequence of operations for T1 and T2 that would have caused this situation. At exactly which point does T2 get blocked when the two transactions execute concurrently?
   (b) Suppose that T2 is making a deposit of 100 € in account 'A-201'. A "SELECT" shows that 'A-201' has a balance of 500€. You could write the subsequent "UPDATE" instruction as "SET balance = 600" or "SET balance = balance + 100". Does it make any difference? Justify your answer.

10. [1.5 pts] Consider a data warehouse of deposits made in bank accounts.
   (a) Write an SQL query to show the total amount of deposits made by year, by customer city, and by branch city.
   (b) Change the previous query: drill-down to month and slice on customer city in order to only show the results for customers living in Lisbon.



Sistemas de Informação e Bases de Dados

2016-01-30

Exam

You may answer the questions in English or Portuguese. Duration: 90 minutes

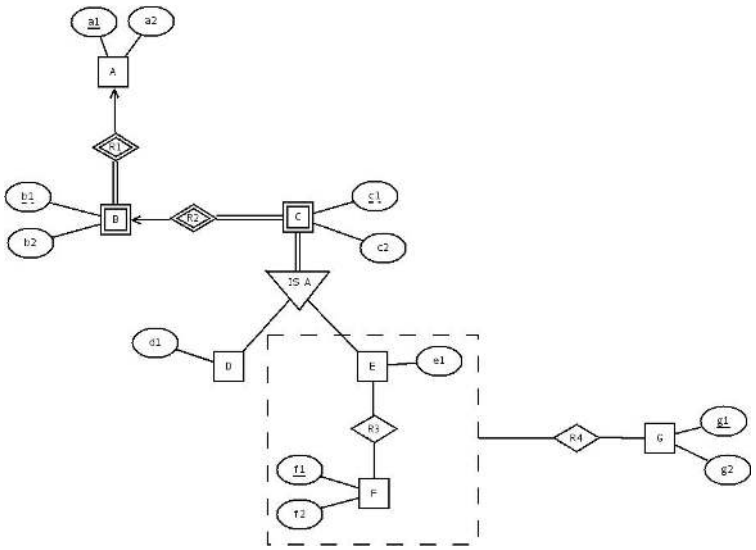
Part I: E-R modeling

You have been hired to design a database for an automobile repair shop. For this purpose, you are given the following description:

Vrum Limited is a multi-brand car servicing station. For every service, the database should store information about the car, namely the registration identifier, model, type of fuel used, and mileage recorded. In addition, every car has at least one driver, but a driver may have several cars. Each driver has a national ID, an address and a phone number. An appointment has to be scheduled for each car that needs to be serviced, and each appointment has a day and a time. It is important to note that while an appointment is always related to one car, a car may have more the one appointment scheduled with the servicing station. After completing the service of an appointment, a repair report should describe the work done; specifically, a report should have a list of repairs, where each repair specifies the part supplied, the cost of the part, and labor charges.

Answer the following questions in a separate sheet of paper

1. [2.5 pts] Draw an E-R diagram for this scenario. Include appropriate constraints in each association.
2. [2.0 pts] The following requirements have to be included in the database design:
   All the mechanics have an employment contract with the repair shop. The mechanics have a national id, an address and a phone number, while each contract is identified by a number and should include contract terms. Not every mechanic has a full-time contract, thus some of the mechanics have part-time contacts. Mechanics with a full-time contract have a list of benefits (e.g., annual leave, sick leave, and health insurance), while part-time contracts do not come with these benefits. A part-time contract must also specify the number of hours of work per week. The repair shop also keeps track of the mechanics that are allocated to each repair.
   Draw an update the previous E-R diagram incorporating the above requirements.
3. [1.5 pts] Convert the following E-R diagram to the relational model.



Part II: SQL and PHP

Consider a database for managing energy consumption in a large organization. There are energy meters scattered across several campi that record the count of energy units consumed (Kwh) at the end of every hour. There are 2 energy rates, A and B, for two periods of consecutive hours.

Meter

MeterID	Building	Department	Campus
001	NorthTower	EE	Alameda
002	NorthTower	EE	Alameda
003	SouthTower	BioEng	Alameda
101	ISTTagus	EE	Taguspark
102	InformaticaII	CSE	Alameda
201	Area51	EE	Nevada

Rate

RateClass	UnitPrice	StartHour	EndHour
A	0.20	08	22
B	0.10	23	07

Reading

MeterID	Month	Day	Hour	Units
001	12	25	00	1023
002	12	25	00	255
001	12	25	01	1024

Answer the following questions in a separate sheet of paper

4. [4.0 pts] Write an SQL query for each of the following questions:
   a) How many readings above 1000 units have been collected from meter 001 in Month 12?
   b) Which building had more readings in Month 12 at the highest UnitPrice? [ORDER BY clause not allowed]
   c) Which departments have buildings in all campi? [HAVING clause not allowed]

5. [2.0 pts] Write an SQL trigger for table Rate to guarantee that the UnitPrice of Rate B can never more than double the UnitPrice of Rate A.

6. [1.5 pts] Write the HTML the code to render the form of the figure and have a web client invoke the URL when "Go!" is clicked with the data also seen on the figure
(the "frame" enclosing the fields is optional).

http://h.pt/*addMeter.php?B="NorthTower"?D="EE"?C="Alameda*"

7. [1.5 pts] Write the PHP code of script *addMeter.php* to register a new meter in the Meter table of the database, given its location (i.e., attibutes Building, Department and Campus). The new MeterID should be the highest MeterID number in the database + 1. The location data is passed in URLs with the syntax of the URL of the previous question.

Part III: Advanced topics

Answer the following questions in a separate sheet of paper

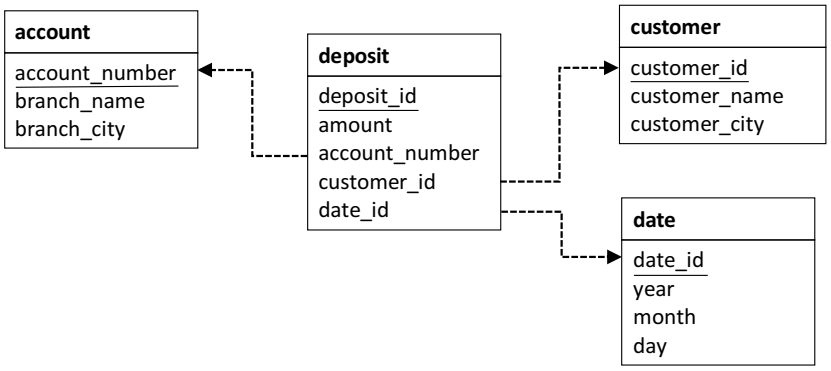
8. [2.0 pts] Consider the following Web page:

```
<form action="deposit.php">
<input type="hidden" name="customer_name" value="<?=$_SESSION['customer_name']?>"/>
<input type="hidden" name="account_number" value="<?=$_REQUEST['account_number']?>"/>
<input type="text" name="amount"/>
<input type="submit"/>
</form>
```

   (a) When the form is submitted, the script deposit.php needs to access the customer name, the account number, and the amount inputs. Which PHP array should the script use to access each of these inputs? Justify your answer.
   (b) Suppose that, in deposit.php, you want to add the account number and the amount to the session data. Write the code to do that.

9. [1.5 pts] Suppose that there are two transactions (T1 and T2) running in parallel in READ COMMITTED mode. At a certain point, T2 gets blocked (i.e. it "freezes") until T1 commits.
   (a) Give a sequence of operations for T1 and T2 that would have caused this situation. At exactly which point does T2 get blocked when the two transactions execute concurrently?
   (b) Suppose that T2 is making a deposit of 100 € in account 'A-201'. A "SELECT" shows that 'A-201' has a balance of 500€. You could write the subsequent "UPDATE" instruction as "SET balance = 600" or "SET balance = balance + 100". Does it make any difference? Justify your answer.

10. [1.5 pts] Consider a data warehouse of deposits made in bank accounts.
   (a) Write an SQL query to show the total amount of deposits made by year, by customer city, and by branch city.
   (b) Change the previous query: drill-down to month and slice on customer city in order to only show the results for customers living in Lisbon.



TÉCNICO LISBOA

Sistemas de Informação e Bases de Dados

03/2/2020

Evaluation Exam (duration: 2h30)

Student Name: _____ **Student Number:** _____

Group 1

An art gallery named "Riscos e Rabiscos S.A." decided to use a relational database to support its day-to-day operations. You were hired you to develop the database!

1a) Represent the following description for the information domain, using an Entity-Association (E-A) diagram:

> *The database must keep information about exhibitions, works of art, and artists.*
>
> *For each exhibition, we need to store a title, the date, a small textual description, and the set of all works of art that were featured in the exhibition (at least one).*
>
> *For the artists, we need to store the name, an address, and the IBAN for his/her bank account. Artists can be uniquely identified by their names, and each has produced at least one work of art.*
>
> *For the works of art, we need to store a title, a price, and a small textual description. Each work of art can only be authored by exactly one artist. The works of art are not uniquely identified by the title, although each artist cannot have produces several works of art with the same title.*
>
> *The art galley may own several works of art that were never featured in an exhibition, aiming to do so only in the future.*

(2v)

1b) Consider the following statements as additions to the previous description for the information domain:

> *The works of art must be categorized as either paintings or sculptures. For sculptures, we need to store the construction materials (maybe more than one). For paintings, we need to store a year. For both paintings and sculptures, we need to store the size (in meters).*
>
> *When works of art are featured in exhibitions, clients may purchase them. We need to store the date for the purchase together with the single corresponding client and, for each client, we need to store a name, an address, and the VAT.*

Represent the corresponding E-A model. It is not necessary to repeat the elements that remain equal to those from the previous question. *(1.5v)*

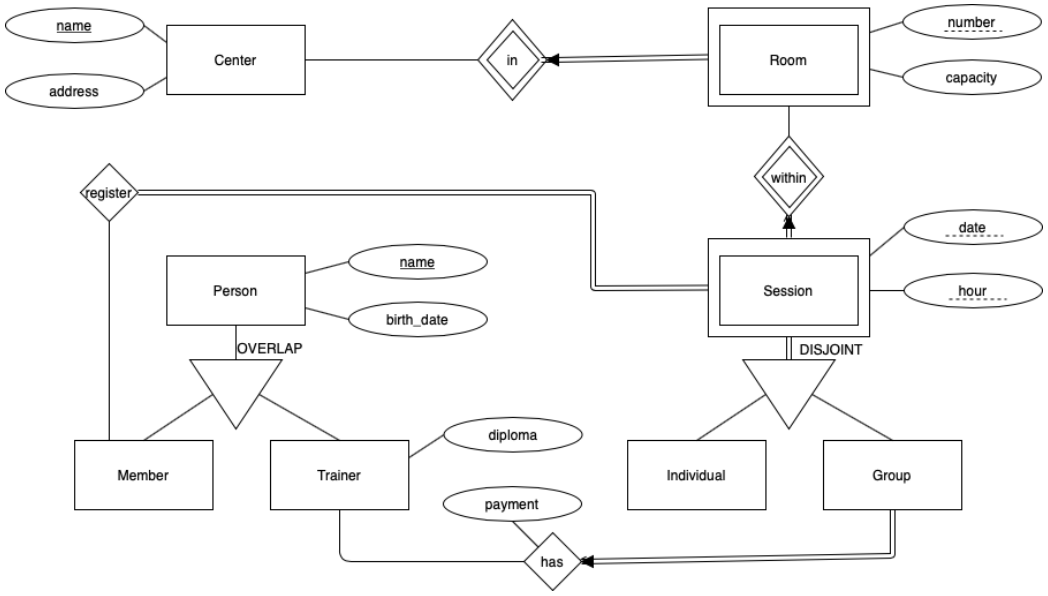
1c) Consider the following statements as additions to the previous description for the information domain:

> *For each exhibition, the art gallery may decide to produce a leaflet. We need to store the text featured in the leaflet, the number of copies that were produced, and the exact works of art featured in the exhibition and whose photos are included in the leaflet.*
> *When clients buy a work of art, a transportation company is always hired to deliver the work to the client. For each transportation company we need to store the corresponding name, its multiple phone numbers, and an insurance number.*

Represent the corresponding E-A model. It is not necessary to repeat the elements that remain equal to those from the previous question. Indicate also, as textual notes, any constraint not captured in the model. *(0.5v)*

Group 2

Convert the following Entity-Association (E-A) model to the relational model. Do not forget to **describe/present integrity constraints for aspects that cannot be captured in the relational model**.



Use the following notation to denote primary keys and foreign keys. *(4v)*

$$relation\_A(\underline{attribute\_1}, attribute\_2, ...)$$
$$attribute\_2 : FK(relation\_B)$$

Group 3

In a database, the relation `City(name,county,country,population,level)` stores data about population and pollution level of cities, the relation `Sensor(code,day,numVehicles,numTrucks)` stores, for each sensor represented by a code, and for each day of the current year, the number of vehicle passages measured by the sensor (trucks and overall), and the relation `Location(sensorCode,cityName)` stores, for each sensor, the city in which it is located.

Present a single SQL statement for each of the following information needs, over the aforementioned database.

3a) Calculate the name, population, and pollution level of the cities in which at least one sensor is located that has detected more than 100 vehicle passages in at least one day. *(2v)*

3b) For each day of the current year and for each city with at least 10,000 inhabitants, calculate the average number of vehicle passages measured in that day by the sensors located in that city. *(1.5v)*

3c) For each city, county, country, and overall, calculate how many truck passages were detected throughout the year by the corresponding sensors. You can use OLAP grouping options within your SQL query. *(0.5v)*

Group 4

Considering the database from the exercises of Group 3, present SQL statements for each of the following questions.

4a) In the context of a single transaction, we aim to perform the following set of operations:

1. Increase by 5% the pollution level of cities where the sensors measure, on average per day, more than 1000 passing vehicles.
2. Insert a new sensor for cities with a pollution level above 10. Initially, the new sensor should not have any measurements recorded in the database.
3. Remove all sensors that measured more than 10000 vehicles in a single day, in cities with a population of less than 1000 individuals.
4. Remove all cities without any sensors.

Show the SQL statements that are involved. *(2v)*

4b) Present a single SQL statement returning the cities for which all the sensors have detected more than 10000 vehicles in a single day. *(1.5v)*

4c) Recall the transaction developed for the first question within this group. Consider that, concurrently to that transaction, another one starts by querying the number of sensors per city, and then the number of sensors per county, in order to compute a statistic from both these values.

When running the second transaction involving two separate queries, on each of the following standard SQL isolation levels, **explain in detail** what are the potential anomalies that can occur over the returned results (i.e., more than naming the anomaly, explain what could actually be the problem with the results). *(0.5v)*

1. Read uncommitted;
2. Repeatable read;

Group 5

Consider the database used in the exercises from Group 3. Answer each of the following questions:

5a) In the relation named "city", consider that there is a functional dependency "county $\mapsto$ country". Assuming that all attribute values are atomic, state in which normal form is the relation and, if needed, present a lossless-join and dependency preserving BCNF decomposition, explaining why these properties hold on the decomposition.

Attend also on the relation named "sensor," which despite being in BCNF could be decomposed into two separate relations: `Sensor1(code,day,numVehicles)` and `Sensor2(code,day,numTrucks)`. Present examples of queries that could benefit from each of the modelling alternatives, explaining your rationale. *(2v)*

5b) Consider a SQL query for retrieving measurements from sensors in Lisbon, for the first 15 days of the year.

Write the query in SQL and state what indexes should be created in order to speed-up its execution, justifying your response. Present also the SQL commands associated with the index creation, and state if the indexes should be based on B+trees or hashing. *(1.5v)*

5c) Consider a relational database management system that supports full-text search leveraging the vector space model with the TF-IDF approach. Explain what are the two key heuristics/components behind the TF-IDF approach for ranking search results for full-text queries. *(0.5v)*

Use this sheet for you drafts. You can separate it from the rest of the test.

–draft–

Sistemas de Informação e Bases de Dados

Exam 2018-01-30

Give your answer directly below each question. You may answer in English or Portuguese. Duration: 90 minutes

Identify all sheets with your name and student number

Number: Name:

Part I: E-R modeling

A hospital aims to reorganize its maternity services. Your role is to create a conceptual model to meet the following description of the maternity sector operation.

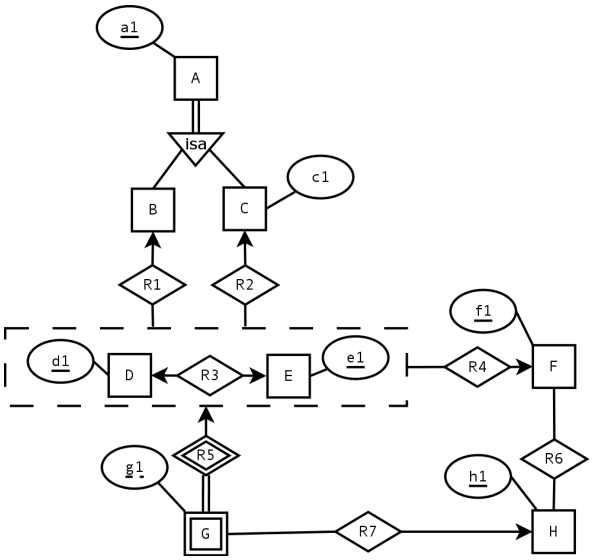
Prior to giving birth, a mother is followed at a maternity of a hospital. Before the first pregnancy, she obtains a unique maternity identification, which is kept together with her citizen card number and birth date. Afterwards, important preparation data is collected during antenatal booking appointments. Every appointment with a mother takes place at a specific date. The preparation data at an appointment includes weight and estimated date of delivery. The mother's final pregnancy step, i.e. labour and delivery, also produces important data, including labour beginning date, time and method used (vaginal, c-section, forceps, etc.). Newborn babies are also registered. Each has a mother and a unique maternity identification, phenotypic sex, date/time of birth, and weigh at birth.

1. [2.0 pts] Draw an E-R diagram for this scenario. Include appropriate constraints in each association and any additional notes if necessary.

2. [2.0 pts] Consider two extensions to the ER model.
    a. *An appointment is characterized not only by the date and mother, but also by the participating obstetrician. Additionally, requests for examination may be produced during an appointment. Each request made in the appointment is sequentially numbered.*
    b. *Obstetricians, mothers and babies are identified by a unified person hospital identification. It should be enforced that a woman must neither be mother of herself, nor simultaneously mother and obstetrician in the same appointment.*

Modify the previous E-R diagram to incorporate the above extensions. You only need to draw the changed and new elements.

3. [1.5 pts] Convert the following E-R diagram to the relational model.



Number: Name:

Part II: SQL and PHP

Consider the following schema of a prescription data warehouse:

Date (<u>dateid</u>, year, month, weekday)

Pharmacy (<u>phid</u>, region, municipality)

Substance (<u>code</u>, name, price_class)

Prescription (<u>pid</u>, dateid, user, phid, sub)
dateid: FK(Date)
phid: FK(Pharmacy)
sub: FK(Substance)

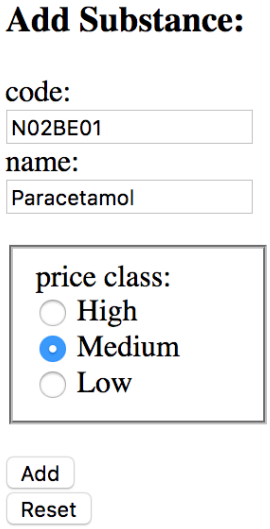
4. [3.0 pts] Write an SQL query for each of the following questions:
   a. How many prescriptions has each user in the database?
   b. What are the pharmacies in every region, excluding the 'Algarve' and 'North' regions, which have sold most prescriptions of "Paracetamol" in December 2017?
   c. Without using the HAVING clause: Which pharmacies have prescriptions of substances that have also been prescribed in all regions where 'Paracetamol' has not been prescribed?

5. [2 pts] The following function returns the number of prescriptions at a municipality on a given year:

   PRESCRIPTIONS (year, municipality)

   Using this function, write a query to obtain the total number of prescriptions in 2017 of the municipalities of the "Algarve" region where the number of prescriptions in 2016 was higher than in 2017.

6. [1.5 pts] Write the code of a HTML form to register a new substance. The design of the form should produce the layout of the figure. When the form is submitted, the script "addSubstance.php" is invoked. The names of the form inputs must be identical to the SQL table column names they correspond to. Also, upon submission, the form data should not be passed in the URL.



7. [1.5 pt] Write the PHP code of addSubstance.php to create a new substance in the database, upon receiving the data submitted through the form. The substance "name" must not exist already in table "Substance". Make sure your code prevents SQL injection.

Number: Name:

Part III: Advanced topics

Consider the data warehouse of the previous chapter:

Date (dateid, year, month, weekday)

Pharmacy (phid, region, municipality)

Substance (code, name, generic, price_class)

Prescription (pid, dateid, user, phid, sub)
dateid: FK(Date)
phid: FK(Pharmacy)
sub: FK(Substance)

8. [1.5 pts] The authorities consider that situations where a user is prescribed more than 10 times in the same day are worth investigating. Write an insert trigger to check if a user should be investigated. If so, then the trigger should generate an exception and output a message with text "START INVESTIGATION".

9. [1.5 pt] The *pid* in table *prescriptions* should be a sequential number, and *auto_increment* is not available in the DBMS.
    a. Explain how each transaction can generate the next *pid*
    b. What is the lowest isolation level that should be used to ensure that upon insertion to table prescription no transaction will attempt to use the same *pid* of another transaction? Explain what would happen if a lower/higer level were selected.

10. [2.0 pts] All PKs and FKs are indexed. For each operation below:
    a) Write an SQL query to obtain the pid of every prescription of substance “N02BE01” to user “U999” and indicate which additional index is required to execute the query using only the index.
    b) Write an SQL query to obtain the distinct substance codes of every prescription in the municipality of “Sintra” and indicate which additional indexes would be required to execute the query without accessing any tables.
    c) Write an SQL query to update the user “U999” in table Prescription to “U001”. Would an index make the query faster? It not, why? If yes, which one?

11. [1.5 pts] The authorities want to analyze prescriptions by date, pharmacy and medication.
    a) Write a query to show the number of prescriptions by year and municipality of 'Paracetamol'.
    b) How would you modify the query above to drill-down on the date dimension to year, month and weekday and print the totals per month and per year?

Sistemas de Informação e Bases de Dados

30/1/2019

Evaluation Exam (duration: 2h30)

Student Name: _____ **Student Number:** _____

Group 1

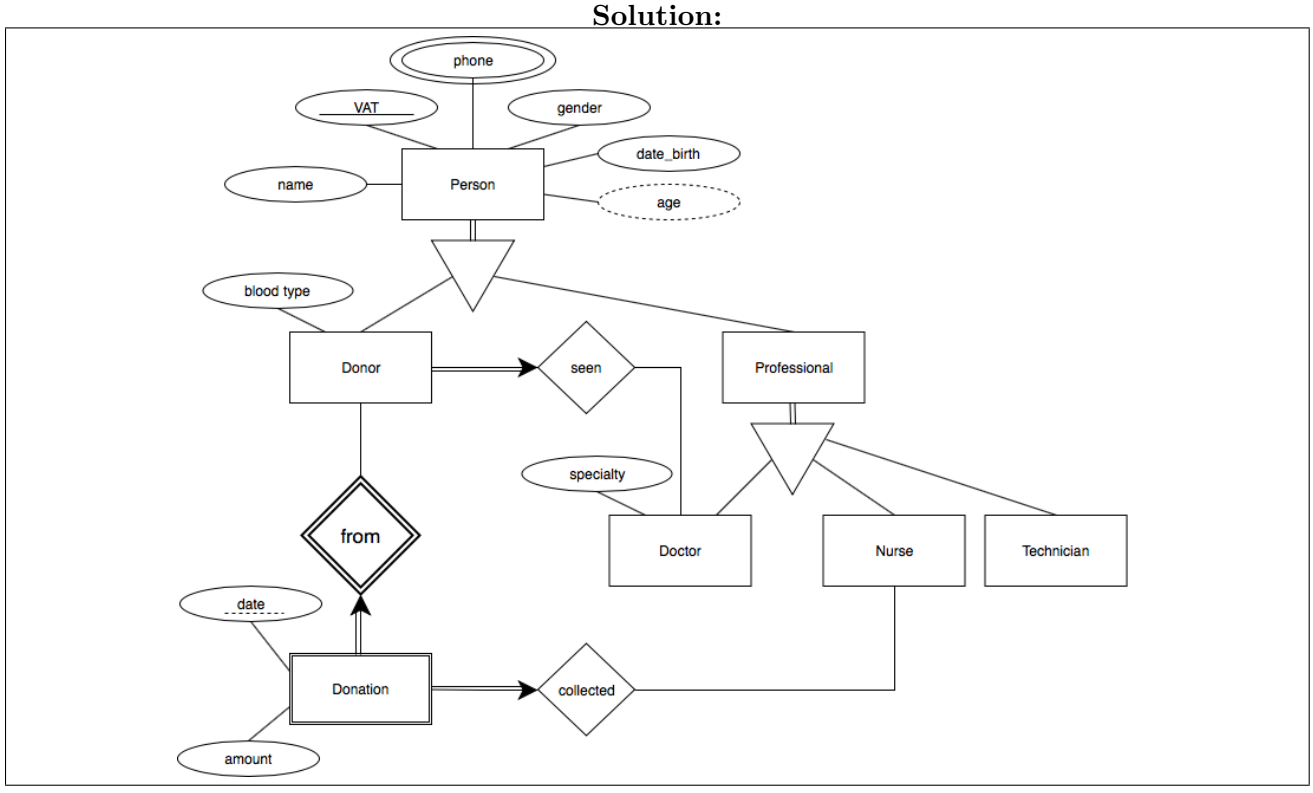
A private blood bank named "Count Duckula S.A." decided on the use of a relational database to support its day-to-day operations. Igor, the company manager, hired you to develop the database!

1a) Represent the following description for the information domain, using an Entity-Association (E-A) diagram:

> *The database must keep information about donors, blood donations, and health professionals working on the blood bank.*
> *For health professionals, we need to store the VAT, name, gender, date of birth, age, and phone numbers (i.e., possibly more than one). The health professionals can be nurses, medical doctors or technicians and, in the case of medical doctors, we need to store the specialty.*
> *For donors, we need to store the blood type, VAT, name, gender, date of birth, age, and phone numbers (i.e., possibly more than one). When donating blood for the first time, donors are examined by one of the medical doctors and registered on the database.*
> *Each blood donation was collected from a single donor at a given date. We need to store information on the amount of blood that was collected, and on who was the nurse performing the collection. Multiple donations can be collected at the same date, although only from different individuals.*

(2v)

Solution:

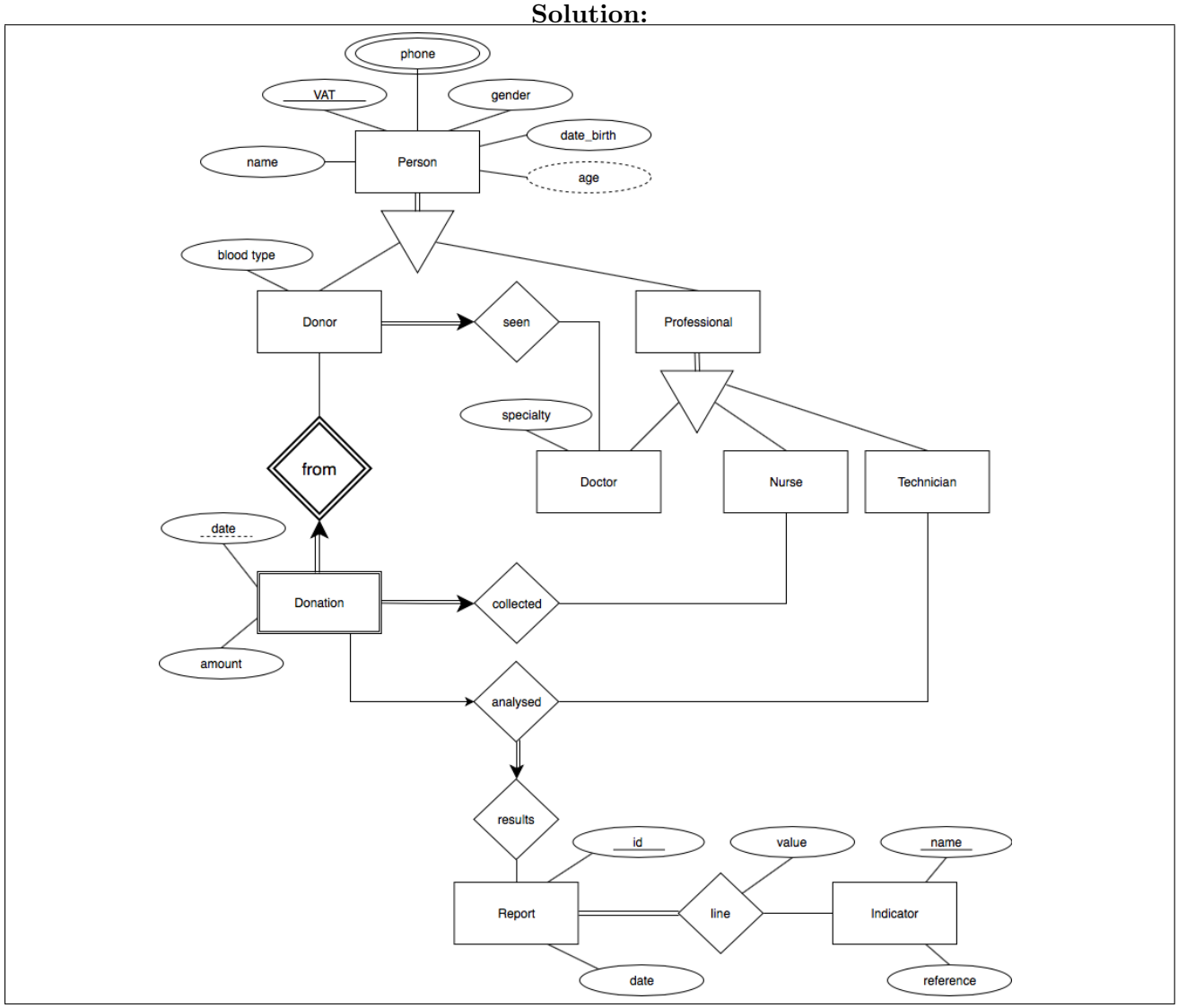


1b) Consider the following statements as additions to the previous description for the information domain:

> *The blood from each donation is analized by a technician, resulting in a report (i.e., a clinical blood panel) describing physiological and biochemical properties. Each of these reports has a unique identifier, a date, and a set of "report lines" describing indicators, values measured for these indicators, and the expected reference values for these indicators. The reference values for a given indicator should be consistent (i.e., the same) across all of the reports.*

Represent the corresponding E-A model. It is not necessary to repeat the elements that remain equal to those from the previous question. *(1.5v)*

Solution:

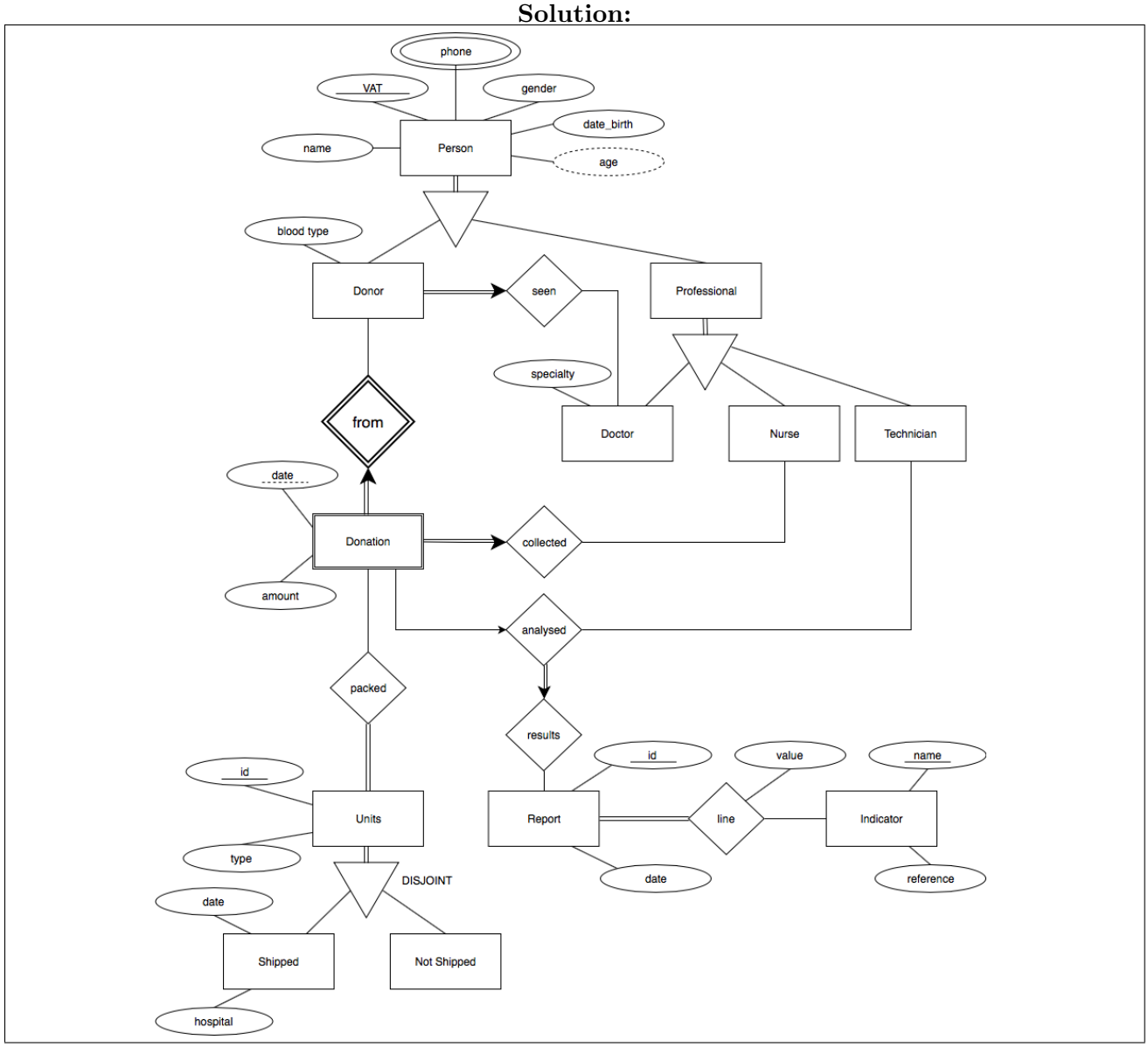


1c) Consider the following statements as additions to the previous description for the information domain:

> *Blood donations are spun in centrifuges to separate them into transfusable components (e.g., red cells, platelets, etc.). Each component is packaged as a "unit," i.e. a standardized amount that doctors will use when transfusing a patient. The database stores information on each of these "units," including a unique identifier, the type of component, and the source donation(s).*
>
> *The units are available to be shipped to hospitals or clinics 24 hours a day, 7 days a week. For each unit that is shipped, the database stores the date and the name of the hospital to where the unit was sent. Multiple units are shipped every day.*

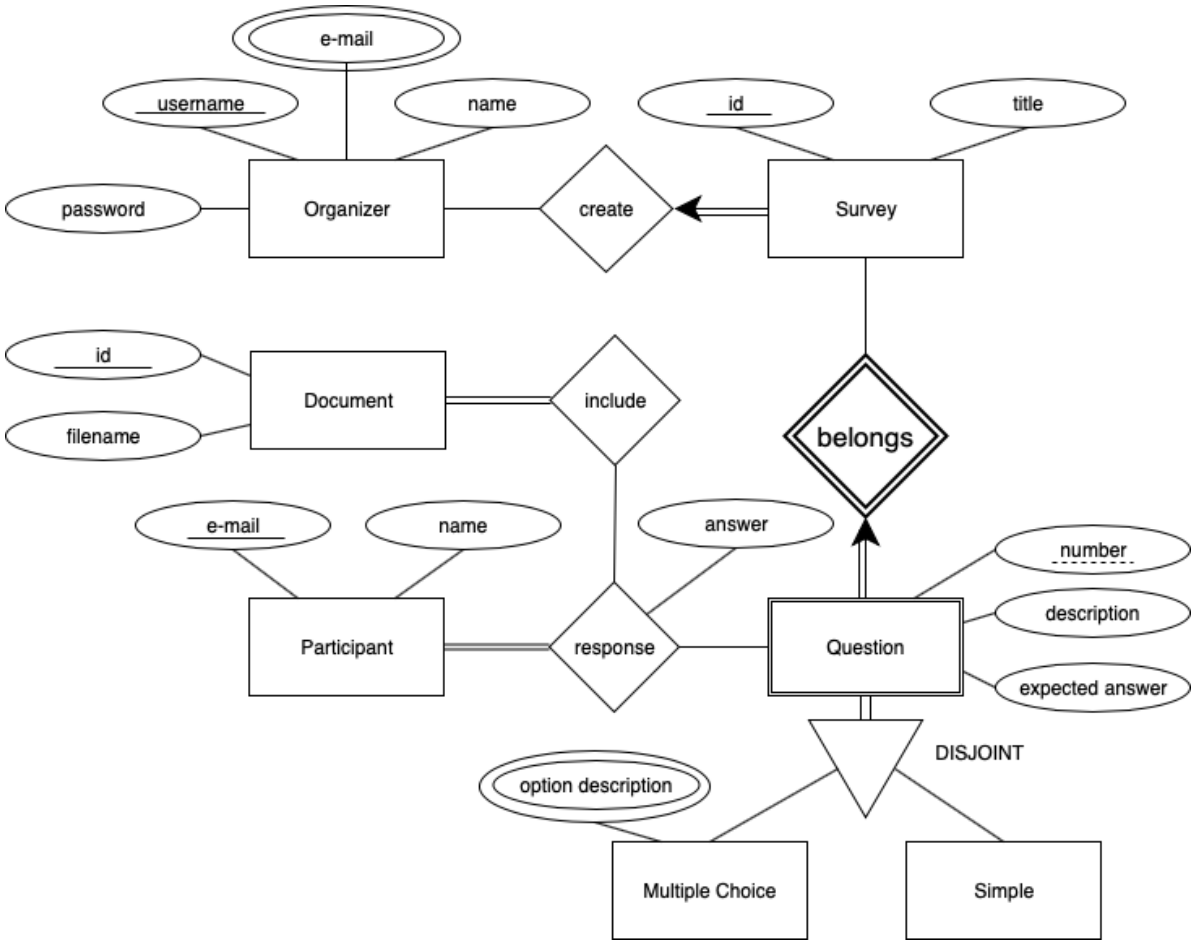
Represent the corresponding E-A model. It is not necessary to repeat the elements that remain equal to those from the previous question. *(0.5v)*

Solution:



Group 2

Convert the following Entity-Association (E-A) model to the relational model. Do not forget to **describe/present integrity constraints for aspects that cannot be captured in the relational model**.



Use the following notation to denote primary keys and foreign keys. *(4v)*

$relation\_A(\underline{attribute\_1}, attribute\_2, ...)$
$attribute\_2 : FK(relation\_B)$

Solution:

$Organizer(\underline{username}, password, name)$
IC: All organizers need to have an email address.

$Email(\underline{username, email})$
$username : FK(Organizer)$

$Survey(\underline{id}, title, username)$
$username : FK(Organizer)$

$Question(\underline{id, number}, description, expected\_answer)$
$id : FK(Survey)$
IC: All questions are either "simple"or "multiple choice", and they cannot be both options simultaneously.

$Simple(\underline{id, number})$
$id, number : FK(Question)$

$Multiple\_Choice(\underline{id, number})$
$id, number : FK(Question)$
IC: All multiple choice questions need to have options.

$Multiple\_Choice\_Option(\underline{id, number, option\_description})$
$id, number : FK(MultipleChoice)$

$Participant(\underline{email}, name)$
IC: All participants reply to at least one question.

$Response(\underline{email, id, number}, answer)$
$email : FK(Participant)$
$id, number : FK(Question)$

$Document(\underline{id}, filename)$
IC: All documents are associated to at least one response.

$Include(\underline{email, id, number, id\_document})$
$email, id, number : FK(Response)$
$id\_document : FK(Document)$

Group 3

Consider that the "Atchoo" clinic relies on a relational database to store information about its activities, using the following relations as part of the database schema.

```
treatments ( disease , medication )

doctors ( name , disease_of_specialization , salary )

treated ( doctor_name , patient_name , diagnostic , treatment , date )
  doctor_name: FK(doctors)
  diagnostic, treatment: FK(treatments)
```

Present a single SQL statement for each of the following information needs:

3a) List the patients who consulted two or more medical doctors, and that were given the same diagnosis by both doctors. List each patient's name and the diagnosis. *(2v)*

Solution:

```
SELECT DISTINCT t1.patient_name, t1.diagnosis
FROM treated t1, treated t2
WHERE t1.patient_name = t2.patient_name
 AND t1.doctor_name <> t2.doctor_name AND t1.diagnostic=t2.diagnostic
```

3b) List the name of the medication(s) that can treat more diseases. *(1.5v)*

Solution:

```
SELECT medication
FROM treatments
GROUP BY medication HAVING COUNT(disease) >= ALL (
 SELECT count(disease) FROM treatments GROUP BY medication )
```

3c) List the disease(s) for which there is only one single medication. *(0.5v)*

Solution:

```
SELECT DISTINCT disease FROM treatments
EXCEPT
SELECT t1.disease
FROM treatments t1, treatments t2
WHERE t1.disease=t2.disease AND t1.medication <> t2.medication
```

Group 4

Considering the database from the exercises of Group 3, present SQL statements for each of the following questions.

4a) In the context of a single transaction, we aim to perform the following set of operations:

1. Remove all information from the "treated" relation on patients diagnosed with "Werewolf Syndrome".
2. Decrease the salary of all doctors that made the previous diagnoses by 30%.

Show the SQL statements that are involved. *(2v)*

Solution:

```
BEGIN TRANSACTION;

UPDATE doctors SET salary = salary - ( salary * 0.3 ) WHERE name IN (
SELECT doctor_name FROM treated WHERE diagnostic = 'Werewolf Syndrome' );

DELETE FROM treated
WHERE diagnostic = 'Werewolf Syndrome';

COMMIT;
```

4b) Present a single SQL query that, using the OLAP grouping options introduced in the classes, can count the number of patients that where treated for each combination of diagnostic and treatment, for each diagnostic, with each treatment, and in total. *(1.5v)*

Solution:

```
SELECT COUNT(DISTINCT patient_name)
FROM treated
GROUP BY CUBE (diagnostic , treatment)
```

4c) Create a trigger to prevent the insertion of medications, in the "treatments" relation, for treating the diseases named "Inflated Ego" and "Verbal Diarrhoea". *(0.5v)*

Solution:

```
CREATE TRIGGER prevent BEFORE INSERT ON treatments
FOR EACH ROW
BEGIN
  IF NEW.disease = 'Inflated Ego' OR NEW.disease = 'Verbal Diarrhoea' THEN
   SET NEW.disease = NULL;
   SET NEW.medication = NULL;
  END IF;
END;
```

Group 5

Consider the database used in the exercises from Group 3. Answer each of the following questions:

5a) In the relation named "treated", consider that there is a functional dependency "diagnostic $\mapsto$ treatment". Assuming that all attribute values are atomic, state in which normal form is the relation named "treated" and, if needed, present a lossless-join and dependency preserving BCNF decomposition. *(2v)*

Solution:

There's a single candidate key and three key attributes, namely 'doctor_name', 'patient_name', 'diagnostic'. Notice that since 'diagnostic' implies 'treatment', the aforementioned three attributes are enough to imply all attributes in the relation.

The relation is in 1NF because all attribute values are atomic.
The relation is not in 2NF because one of the non-key attributes (i.e., 'treatment') depends only on part of the candidate key (i.e., 'diagnostic').

The relation is not in 3NF because it is also not in the 2NF.
A lossless BCNF decomposition would be:

```
treated ( doctor_name , patient_name , diagnostic , date )
treated_procedure ( diagnostic , treatment )
```

The decomposition is lossless-join because the common attribute in the new relations is a candidate key in one of the new relations.

5b) Consider SQL queries for retrieving (a) the doctors specializing in "Verbal Diarrhoea", and (b) the doctors with a salary above 10000 Euros. For each query, describe an index that could improve its performance. Display the SQL commands associated with the index creation, and justify your response, stating if the indexes should be based on B+trees or hashing. *(1.5v)*

Solution:

For the first case we should use an hashing index on the attribute disease_of_specialization, given that hash indexes have a better performance for equality queries, particularly when considering attributes taking strings as values.

```
CREATE INDEX idx1 ON doctors (disease_of_specialization) USING HASH;
```

The second case would require an index based on B+trees (perhaps a secondary index, given that we expect to retrieve only few records).:

```
CREATE INDEX idx2 ON doctors (salary) USING BTREE;
```

5c) Explain what are "SQL injection flaws" and how they can be prevented in PHP Web applications. *(0.5v)*

Solution:

SQL Injection flaws are introduced when software developers create dynamic database queries that include user supplied input, allowing an attacker to inject new SQL code into the query that would be executed by the database. The use of prepared statements with variable binding (aka parameterized queries) prevents SQL injection attacks, ensuring that an attacker is not able to change the intent of a query.

Use this sheet for you drafts. You can separate it from the rest of the test.

–draft–

Sistemas de Informação e Bases de Dados

Exame de 2ª Época

This exam is written in English, but you may answer in Portuguese.

Introduction – Consider the following scenario:
An online supermarket allows customers to buy products and have those products delivered to a certain address. Each customer has a name, username and password. Each product has a barcode, description, and price. A shopping basket has a unique id, and it may contain multiple products, with a certain quantity of each product (e.g. 2 bottles of juice, 6 yogurts, 1 toothpaste, etc.).
To complete the purchase, the customer specifies an address for delivery, and a credit card for payment. Customers may have multiple addresses and multiple credit cards, and it is necessary to store all this information in the database. Each address has a street and a zip code. Each card has a title holder, a card number, and an expiry date.

An IT company was hired to create a database for this online supermarket. Here is their solution:

customer(name, <u>username</u>, password)

address(<u>username, street, zipcode</u>)
username : FK(customer)

creditcard(holder, <u>number</u>, date, username)
username : FK(customer)

product(<u>barcode</u>, description, price)

basket(<u>id</u>, username, street, zipcode, number)
username, street, zipcode : FK(address)
number : FK(creditcard)

contains(<u>id, barcode</u>, quantity)
id : FK(basket)
barcode : FK(product)

Part I: E-R **[Answer in a separate sheet of paper]**

1. From the solution proposed by the IT company, draw an E-R (Entity-Relationship) diagram that captures their conceptual view of this scenario.
   a) First, draw the part that involves *customer*, *address*, and *creditcard*.
   b) Complete the diagram with the part that involves *product*, *basket*, etc.
   (You can draw both parts in the same diagram.)

2. If customers were allowed to have only one address and only one credit card, how would you change the database design?
   a) Draw a new E-R diagram that takes advantage of this simplification.
   b) Indicate which database tables would be changed, and how they would be changed.

Part II: SQL **[Answer in a separate sheet of paper]**

3. Assume that you have the database as originally designed by the IT company. Write an SQL query to answer each of the following questions:
   a) What is the most expensive product in the supermarket?
   b) Who is the customer who has the most expensive basket? (in terms of the total price of the products contained in the basket)

c) How many customers have only one address and only one credit card in the database?
d) Suppose that the description of each product is a string that contains the brand (e.g. Danone), the type of product (e.g. yogurt), and possibly other data such a flavor (e.g. banana), size (e.g. 160 gr), etc.. A complete example is: 'Danone yogurt banana (160 gr)'.
The question is: how many products does Danone sell which are not yogurts?

4. A basket is associated with a customer in two different ways: through the address, and through the credit card. (See the foreign keys from *basket* to other tables in the database.)
   a) Is this correct/appropriate? Or could this make the database become inconsistent? Justify your answer and explain your view on this topic.
   b) To avoid credit card frauds, the supermarket requires that, when creating a basket, the card holder (*holder* in table *creditcard*) must be equal to the customer name (*name* in table *customer*). Write a database trigger to ensure that this is always the case.

Part III: PHP [Answer in a separate sheet of paper]

5. Suppose that you have to develop an HTML form and a PHP script to add products to a basket. (Assume that the products and the basket already exist in the database.)
   a) The user must be able to choose multiple products and choose the basket they will be added to. What kind of HTML elements would you use in the HTML form? Draw the HTML form and explain its elements. (Use only HTML elements that you have learned in this course.)
   b) Write a sketch of the PHP script that receives the data from the form and adds the product to the basket. In particular, show the PHP code/instructions and the SQL query.

Part IV: Advanced Topics [Answer in a separate sheet of paper]

6. There is a new table to store the quantity of each product that is available in stock:

stock(<u>barcode</u>, quantity)
barcode : FK(product)

   a) We need to be able to quickly find those products which are sold out (i.e. whose quantity in stock is zero). How would you use indexes for this purpose? Explain your answer and present the necessary SQL code.
   b) Suppose that the values to be indexed are 5, 3, 2, 1, 0 (in this order). The index is a B^+-tree with a maximum of two values in each node. Draw the B^+-tree after the insertion of each value.
   c) Whenever a product is added to a basket (see table *contains*) the same quantity must be subtracted from the *stock* table. How would you use transactions to ensure that the quantity in stock never becomes negative? Explain what the transaction would do. Also, give an example of how the quantity in the stock could become negative without the use of transactions.

7. In an execution plan, there are several algorithms to do the same thing. Why? Isn't it enough to have just one? Explain your answer, and give examples of algorithms that you know.

8. When the database system crashes, it applies *undo* to some transactions and *redo* to others. How does the system know what to *undo* or *redo*? And what exactly does it mean to *undo* and to *redo* a transaction? Explain your answers in detail.

Sistemas de Informação e Bases de Dados

Resolução de exercícios de exame

Matéria

- Modelo E-A e conversão
- Consultas em SQL
- Funções e triggers
- Desenvolvimento de aplicações com PHP
- Índices e planos de execução
- Transacções e recuperação

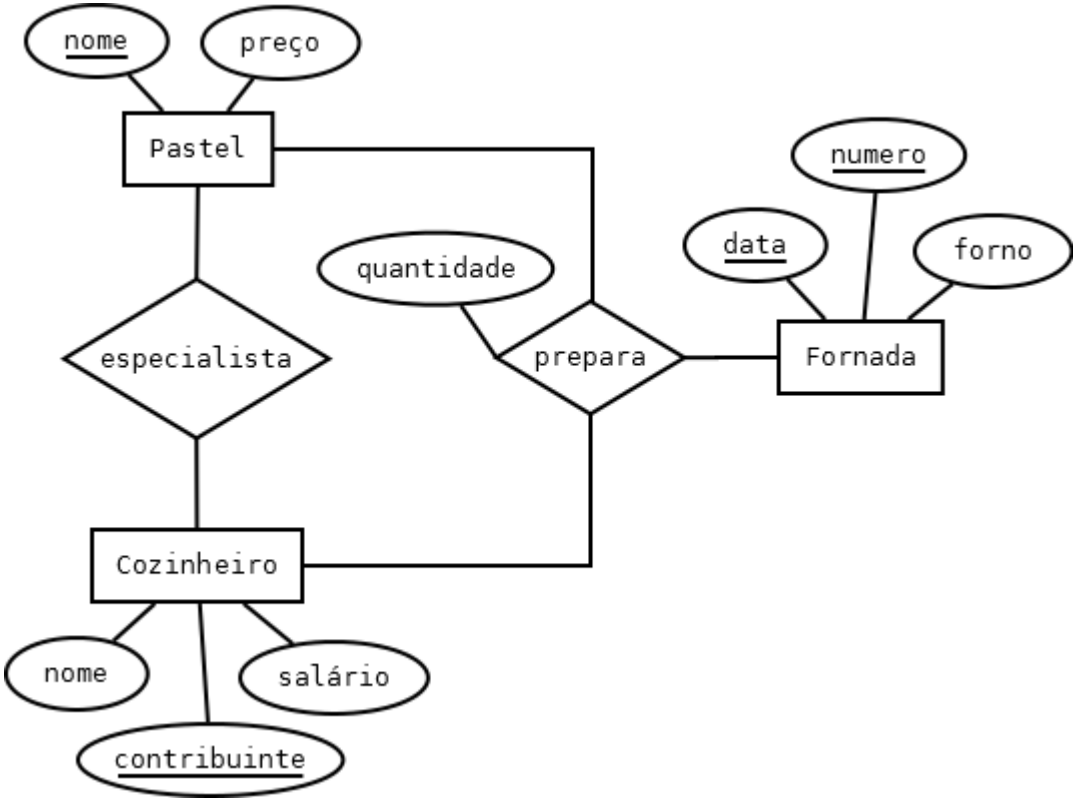
Exercícios

- Modelo E-A

Uma pastelaria vende vários tipos de doces e pastéis. Cada pastel é identificado por um nome ("pastel de nata", "bola de berlim", etc.) e tem um preço. A pastelaria tem vários cozinheiros, e cada cozinheiro é especialista na confecção de um ou mais pastéis. Cada cozinheiro tem um nome, número de contribuinte e salário. Os cozinheiros preparam várias fornadas de cada pastel por dia. Cada fornada tem uma data (e.g. 2012-01-09), um número (1, 2, 3, ...) que se reinicia todos os dias, e o forno em que foi produzida. Em cada fornada, um cozinheiro prepara uma certa quantidade de um pastel.

Exercícios

- Modelo E-A



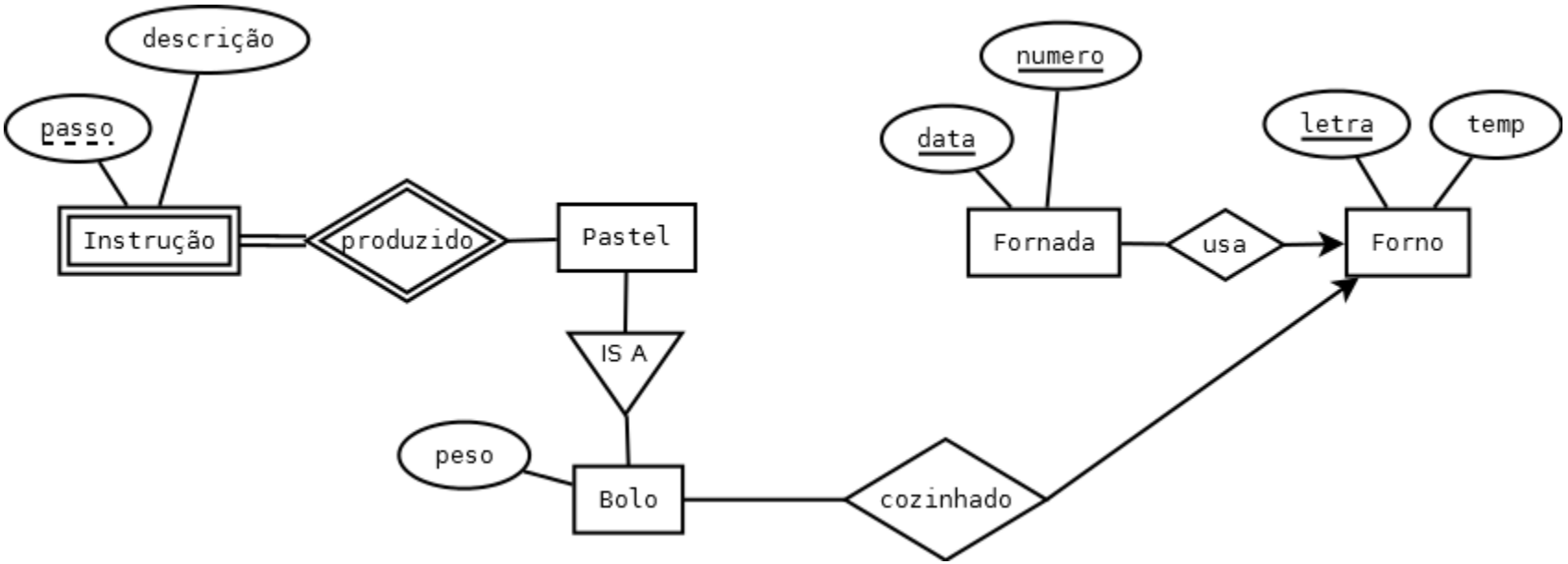
Exercícios

- Modelo E-A

Cada forno é identificado por uma letra (A, B, C, ...) e tem uma temperatura máxima. Cada pastel deve ser produzido de acordo com um conjunto de instruções (por exemplo: no passo 1 prepara-se a massa com farinha, no passo 2 junta-se os frutos secos, etc.). Cada instrução têm um número sequencial (passo) e uma descrição. Para além de pastéis, a pastelaria também vende bolos de aniversário. Os bolos são semelhantes aos pastéis em vários aspectos: são identificados por um nome; têm um preço; e também são produzidos através de um conjunto de instruções. As diferenças face aos pastéis é que um bolo tem um peso e tem um forno específico onde pode ser cozinhado, não podendo ser cozinhado noutro forno a não ser esse.

Exercícios

- Modelo E-A



Exercícios

- Conversão

Converta o diagrama resultante para modelo relacional. Use a seguinte notação para indicar as chaves primárias (sublinhado) e estrangeiras (FK):

$$relacao_A(\underline{atributo_1}, atributo_2, ...) \qquad atributo_2 : FK(relacao_B)$$

Exercícios

- Conversão

$Pastel(\underline{nome}, preco)$

$Cozinheiro(\underline{contribuinte}, nome, salario)$

$Fornada(\underline{data, numero}, letra)$

$letra : FK(forno)$

$especialista(\underline{nome, contribuinte})$

$nome : FK(Pastel)$

$contribuinte : FK(Cozinheiro)$

$prepara(\underline{nome, contribuinte, data, numero}, quantidade)$

$nome : FK(Pastel)$

$contribuinte : FK(Cozinheiro)$

$data, numero : FK(Fornada)$

$Forno(\underline{letra}, temp)$

$Instrucao(\underline{nome, passo}, ingred, quant)$

$nome : FK(Pastel)$

$Bolo(\underline{nome}, peso, letra)$

$nome : FK(Pastel)$

$letra : FK(forno)$

Exercícios

- SQL

Considere uma base de dados semelhante à do exercício anterior, em que se guardam dados sobre pastéis e cozinheiros, bem como sobre os pastéis que cada cozinheiro sabe preparar.

Tabela *pastel*

designação	preço
pastel de nata	0.80
bola de berlim	1.20
...	...

Tabela *cozinheiro*

contribuinte	nome	salário
210210	João	500.00
312312	Pedro	600.00
456456	Susana	550.00

Tabela *prepara*

designação	contribuinte
pastel de nata	210210
pastel de nata	312312
bola de berlim	312312

1.

Quem são os cozinheiros que só sabem fazer pastéis de nata? Na resposta, identifique o cozinheiro pelo número de contribuinte.

Exercícios

- SQL

```
select contribuinte
from prepara as p1
where p1.designacao = 'pastel de nata'
and not exists (select designacao
                from prepara as p2
                where p2.contribuinte = p1.contribuinte
                and designacao <> 'pastel de nata')
```

Exercícios

- SQL

Considere uma base de dados semelhante à do exercício anterior, em que se guardam dados sobre pastéis e cozinheiros, bem como sobre os pastéis que cada cozinheiro sabe preparar.

Tabela *pastel*

designação	preço
pastel de nata	0.80
bola de berlim	1.20
...	...

Tabela *cozinheiro*

contribuinte	nome	salário
210210	João	500.00
312312	Pedro	600.00
456456	Susana	550.00

Tabela *prepara*

designação	contribuinte
pastel de nata	210210
pastel de nata	312312
bola de berlim	312312

2.

Quem é o cozinheiro que sabe preparar mais pastéis? Na resposta, indique o contribuinte e o número de pastéis que o cozinheiro sabe preparar.

Exercícios

- SQL

```
select contribuinte, count(designacao)
from prepara
group by contribuinte
having count(designacao) >= all(select count(designacao)
                                from prepara
                                group by contribuinte)
```

Exercícios

- SQL

Considere uma base de dados semelhante à do exercício anterior, em que se guardam dados sobre pastéis e cozinheiros, bem como sobre os pastéis que cada cozinheiro sabe preparar.

Tabela *pastel*

designação	preço
pastel de nata	0.80
bola de berlim	1.20
...	...

Tabela *cozinheiro*

contribuinte	nome	salário
210210	João	500.00
312312	Pedro	600.00
456456	Susana	550.00

Tabela *prepara*

designação	contribuinte
pastel de nata	210210
pastel de nata	312312
bola de berlim	312312

3.

Se todos os cozinheiros com nome terminado em 'o' fizerem greve, quais são os pastéis que a pastelaria pode fazer nesse dia? Na resposta, não apresente duplicados.

Exercícios

- SQL

```
select distinct(designacao)
from prepara natural join cozinheiro
where nome not like '%o'
```

Exercícios

- Funções

Considere o seguinte modelo relacional

piloto(<u>npiloto</u>, nacionalidade)
etapa(<u>netapa</u>, kms, piso)
percorre(<u>netapa, npiloto</u>, tempo)
netapa : FK(etapa) npiloto : FK(conduz)

Escreva uma função SQL que devolve o piloto vencedor de uma dada etapa (menor tempo na etapa).

Exercícios

- Funções

```
create function vencedor_etapa (varchar(255)) returns varchar(255)
as
$$
    declare vencedor varchar(255);
    begin
        select npiloto into vencedor
        from percorre
        where netapa = $1
            and tempo <= all (select tempo
                              from percorre
                              where netapa = $1)
        return vencedor;
    end
$$
language plpgsql;
```

Exercícios

- Triggers

Considere o seguinte modelo relacional

carro(matricula, construtor, modelo)
conduz(npiloto, ncopiloto, matricula)
npiloto : FK(piloto) ncopiloto : FK(piloto) matricula : FK(carro)
piloto(npiloto, nacionalidade)

Escreva um trigger para impedir que um par já existente de piloto e co-piloto seja inserido na tabela "conduz" com papéis trocados (o piloto como co-piloto e o co-piloto como piloto). Inclua todas as instruções necessárias para criar o trigger.

Exercícios

- Triggers

```
create function cancel_pilots_proc() returns trigger
as $$
begin
    if exists(select * from conduz where npiloto = new.ncopiloto and ncopiloto = new.npiloto) then
        raise exception 'Erro: o mesmo par de piloto e co-piloto já existe.';
    end if;
    return new;
end
$$ language plpgsql;

create trigger cancel_pilots after insert on conduz
for each row execute procedure cancel_pilots_proc();
```

Exercícios

- PHP

Suponha que quer fazer uma página Web, em PHP, que executa uma consulta sobre uma base de dados PostgreSQL e apresenta os seus resultados. Indique quais são as funções PHP que utilizaria e para que serviriam.

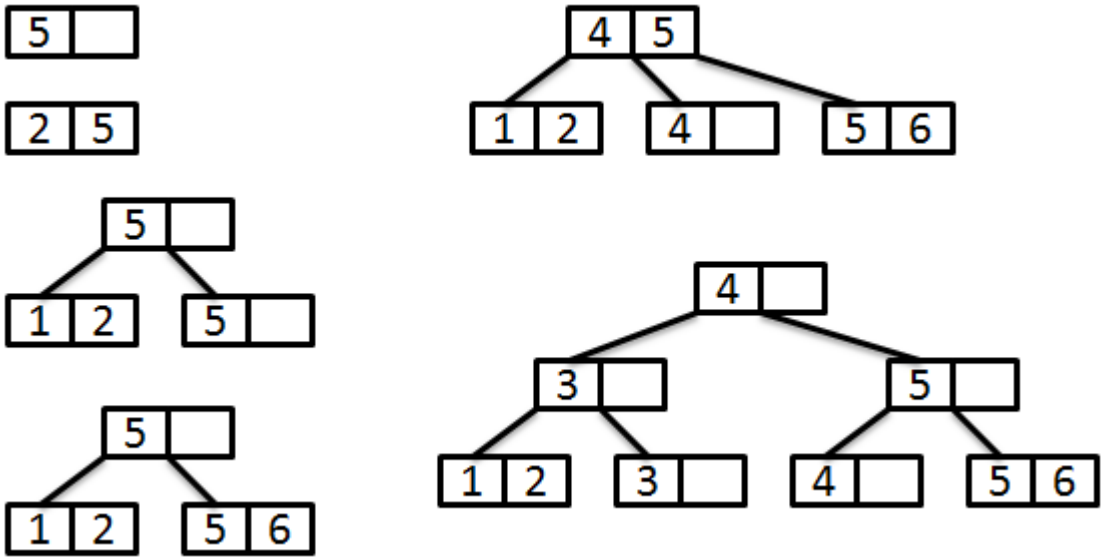
As funções a utilizar seriam:

1. `pg_connect`: abre a ligação à base de dados;
2. `pg_query`: executa a consulta na base de dados;
3. `pg_fetch_assoc`: processa os resultados da consulta executada;
4. `pg_close`: fecha a ligação à base de dados.

Exercícios

- Índices

Considere uma árvore B+ com n=3, inicialmente vazia. Desenhe a árvore passo a passo à medida que insere os elementos 5, 2, 1, 6, 4, 3 por esta ordem.



Exercícios

- Planos de execução

Indique que algoritmos de junção conhece.

- nested-loop join
- merge-join
- hash-join

Exercícios

- Transacções

Na execução de transacções, explique qual é o interesse de haver vários níveis de isolamento possíveis (desde *SERIALIZABLE* a *READ UNCOMMITTED*). O que se ganha em não forçar todas as transações a operar no nível de isolamento mais elevado?

Os vários modos de isolamento têm impacto na concorrência de transacções. O modo SERIALIZABLE é o que impõe maior isolamento, enquanto que o modo READ UNCOMMITTED é o que impõe menos, permitindo que uma transacção aceda a valores escritos por outras transacções que ainda não fizeram commit. Embora por princípio deva ser usado o maior nível de isolamento, este é também aquele que permite menor concorrência. Sendo assim, em certos casos onde não é necessário precisão absoluta (no cenário do banco, p.ex. no caso do cálculo da média dos saldos de todas as contas) poderá ser usado um nível de isolamento mais baixo para permitir que o resultado seja obtido mais rapidamente, sem ter de aguardar pela conclusão das transacções em curso.

Exercícios

- Transacções

Explique sucintamente as quatro propriedades fundamentais das transacções (*ACID*), e diga quais dessas são asseguradas pelos mecanismos de recuperação dos sistemas de bases de dados.

As quatro propriedades são a Atomicidade (ou todas as instruções da transacção ou nenhuma são feitas), a Consistência (antes e depois da transacção, a base de dados deve estar num estado consistente), o Isolamento (enquanto uma transacção não fez commit, outras transacções não devem aceder aos mesmos dados), e Durabilidade (uma vez concluída, o resultado da transacção passa a ser o novo estado da base de dados). Destas, os mecanismos de recuperação permitem garantir a Atomicidade (ou tudo ou nada é feito, através de redo/undo), a Consistência (só há dois estados possiveis, antes e depois da transacção) e a Durabilidade (redo garante que o resultado da transacção fica guardado).

Exercícios

- Recuperação

Considere o seguinte *log* de uma bases de dados

$< T_1\ start >$
$< T_1, A, 20, 10 >$
$< T_1, B, 20, 30 >$
$< T_1\ commit >$
$< T_2\ start >$
$< T_2, B, 30, 25 >$
$< T_3\ start >$
$< T_3, B, 25, 20 >$
$< checkpoint >$
$< T_2, A, 10, 15 >$
$< T_2\ commit >$
$< T_3, C, 20, 25 >$
$\longrightarrow crash$

Indique, pela ordem correcta, as operações de *redo* e *undo* que o sistema terá de fazer. Para cada uma dessas operações, explique por que razão deve ser feito o *redo* ou *undo*.

Exercícios

- Recuperação

Relativamente a T_1 não é necessário fazer nada uma vez que fez *commit* antes do *checkpoint*. T_2 tem de sofrer redo porque o *commit* está no *log*. T_3 tem de sofrer undo porque o *commit* não está no *log*. Primeiro o sistema fará *undo* de $< T_3, C, 20, 25 >$ e de $< T_3, B, 25, 20 >$, seguido de *redo* de $< T_2, A, 10, 15 >$.

Exercícios

- Recuperação

Considere a seguinte execução de 3 transacções:

T1 :	read(A)
T2 :	read(B)
T3 :	read(A)
T3 :	read(B)
T3 :	A = A + 100
T3 :	write(A)
T3 :	B = B - 50
T3 :	write(B)
T1 :	A = A + 100
T1 :	write(A)
T2 :	read(B)
T2 :	B = B - 50
T2 :	write(B)

Durante este escalonamento, o sistema sofreu um *crash* e após reiniciar fez *redo* de T1 e de T3, e fez *undo* de T2. Indique:

- Antes do *crash*, que transacções já tinham feito commit e onde é que esse commit pode ter acontecido? Justifique.
- Assumindo que as transacções fizeram commit o mais cedo possível, onde é que o crash pode ter ocorrido? Justifique.

Exercícios

- Recuperação

Se o sistema fez redo de T1 e T3 é porque estas já tinham feito commit. O commit de T1 só pode ter acontecido entre T1:write(A) e o crash. O commit de T3 só pode ter acontecido entre T3:write(B) e o crash.
Assumindo que T1 e T3 fizeram commit antes de T2:read(B), então o crash pode ter acontecido em qualquer ponto a partir daí.