

INFO-H-415 – Advanced databases

First session examination

The exam is divided in four sections. All sub-questions are worth approximately the same amount of points. However, some of these will only take you a minute, some require a bit more thinking, and a couple would require weeks to answer perfectly. Make the best use of your time.

1 Active Databases (3 pt)

The international union of Bruce Springsteen's fan clubs uses the following relational database:

- **Fan** (*fNo*, *firstName*, *lastName*, *groupNo*, *concerts*, *vip*)
groupNo references **LocalGroup.gNo**
- **LocalClub** (*cNo*, *cName*, *city*, *president*)
president references **Fan.fNo**
- **LocalMeeting** (*clubNo*, *date*)
clubNo references **LocalClub.cNo**
- **GoesToMeeting** (*clubNo*, *date*, *fanNo*)
clubNo references **LocalClub.cNo**
fanNo references **Fan.fNo**
- **Concert** (*date*, *city*)
- **GoesToConcert** (*fanNo*, *date*)
fanNo references **Fan.fNo**

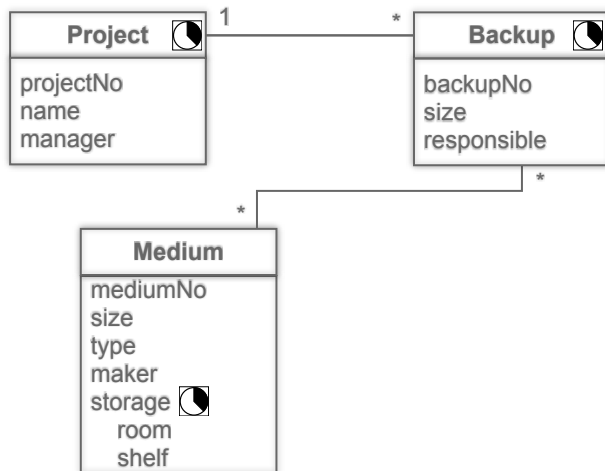
Questions

Write the code ensuring the following **integrity constraints**. Whenever multiple triggers are needed to enforce a single integrity constraint, list all of them, but write the code in full for only one of them. Throughout the entire question you should provide at least one example for each of ON INSERT, ON UPDATE, and ON DELETE triggers.

1. The president of a local club must be a VIP member of that group.
2. There is no local meeting on a date where Bruce Springsteen plays in the city of that local fan club.
3. A fan can only assist to a meeting of a local club he/she belongs to, unless he/she is a VIP. In that latter case, he/she can assist any club meeting.
4. The attribute *concerts* is derived from the entries of table **GoesToConcert**.
5. The attribute *vip* is derived from the entries of table **GoesMeeting**. It is initialised to `False` and set to `True` if the fan has assisted to at least 100 meetings.

2 Temporal Databases (7 pt)

A company manages its IT projects by means of the following database:



Project describes the project, with its name, its manager, and the time interval where this project was active.

Backup describes each completed backup for a given project, the time period it covers, the total size of the backup, and the responsible system engineer who has done the backup.

Medium describes the set of available media (CD-Rom, zip disk, magnetic tapes, etc.) and their size.

MediumStorage indicates where and when the backups have been stored. The company has several rooms (attribute *room*) where numbered shelves (attribute *shelf*) allow an efficient storage.

BackupMedium indicates which medium has been used for which backup.

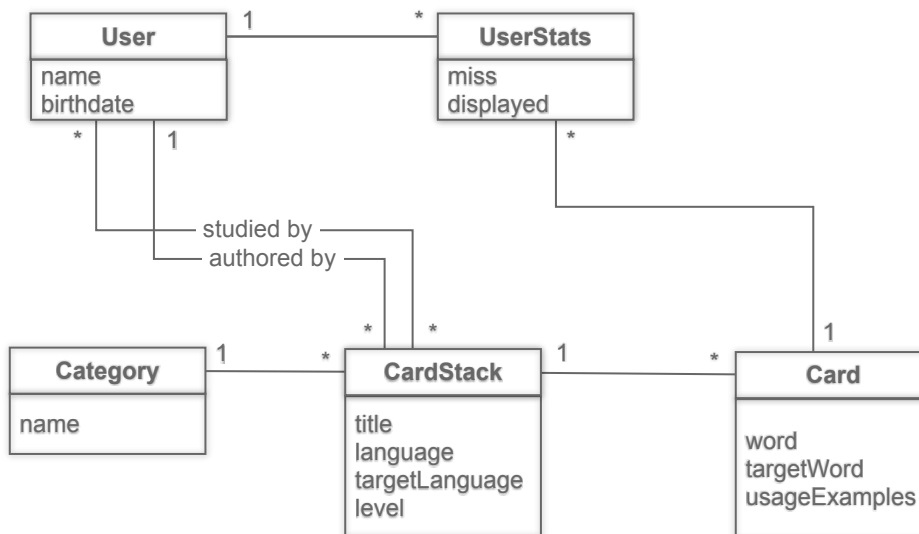
Questions

1. Define a **relational schema** corresponding to the conceptual schema above.
2. Write the code ensuring the following **integrity constraints**.
Whenever multiple triggers are needed to enforce a single integrity constraint, list all of them, but write the code in full for only one of them. Throughout the entire question you should provide at least one example for each of ON INSERT, ON UPDATE, and ON DELETE triggers.
 - (a) Once a first backup has been done of a project, the media type cannot be changed for the subsequent backups of that same project.
 - (b) In each room, there are at most 2000 locations on the shelves.
 - (c) A backup can only be performed after its associated project has begun.
 - (d) The medium's size must be at least as large as the project size it is supposed to backup.
3. Write the following **queries** in SQL

- What is the reference number and the capacity of all the media that were stored on 1st January 2014 in room 42?
- List the name of the closed projects for which no backup has been done after the project was closed, as well as the name of the projects that have not been backed-up during the last two months.
- Give the storage media that have been used to backup two different projects during the same time period.
- Give the project names for which the backup size has only grown. Consider the size at backup time.
- Give the history of the total size of all backups contained in each room.

3 Object Databases (5 pt)

An e-learning company offers its clients sets of flash cards, that allow them to memorise vocabulary for foreign languages. Users can subscribe to different card sets (also called 'stacks'). For each such stack, the user can study the corresponding cards: he or she is shown one card at a time (*displayed* represents the number of times that a given card has been shown to him/her) and asked to select the corresponding translation. If the user fails to give the correct answer, *miss* is incremented for that card. Each stack has a difficulty level defined between 0 and 300.



Here is part of the associated types in Oracle:

```

CREATE TYPE TUser;
CREATE TYPE TUserStats;
CREATE TYPE TCategory;
CREATE TYPE TCardStack;
CREATE TYPE TCard;
CREATE TYPE TSetRefUsers AS TABLE OF REF TUser;
CREATE TYPE TSetRefCards AS TABLE OF REF TCard;
CREATE TYPE TSetRefCardStacks as TABLE OF REF TCardStack;
CREATE OR REPLACE TUser AS OBJECT (
    name VARCHAR2(40),

```

```

    birthdate DATE,
    stats REF TUserStats,
    studied TSetRefCardStacks,
    authored TSetRefCardStacks
);
CREATE OR REPLACE TYPE TUserStats AS OBJECT(
    miss INTEGER,
    displayed INTEGER,
    user REF TUser,
    card TSetRefCards
);

```

Questions

1. Write the **type definitions** for TCardStacks, TCard, and TCategory.
2. Write the following **Oracle queries**:
 - (a) Give the name of all users that have registered to learn more than 500 cards.
 - (b) List all cards, grouped in three difficulty groups: *easy* (difficulty level less than 100), *medium* (level between 100 and 200), and *hard* (level higher than 200)
 - (c) Give the name of the users that have registered to learn all stacks associated to a given category that have been authored by another user.
 - (d) For each user, give: 1) the number of stacks he or she has registered for; 2) for each language, the highest level of stacks he or she has registered for; 3) the total number of studied cards (a card is considered as having been studied if it has been shown more than five times and the hit rate is higher than 80%).

4 Spatial Databases (5 pt)

Consider a database that is used to manage numerical control programs for CNC milling machines. Such machines allow cutting holes and shapes in various materials such as wood or metal sheets and are used extensively in furniture production factories like, for instance at IKEA. Most pieces of furniture are composed of a number of assembled wooden chipboard elements (often several of the same shape). Before being cut, each element is pre-drilled at several coordinates (always with a diameter of 5mm) to allow later assembly. Let us assume that the database used has the following structure:

- **Model** (*mNo*, *name*)
- **Element** (*modelNo*, *shapeNo*, *count*)
modelNo references **Model**.*mNo*
shapeNo references **Shape**.*sNo*
- **Shape** (*sNo*, *geom*)
- **ShapeHole** (*shapeNo*, *seq*, *point*)
shapeNo references **Shape**.*sNo*

A furniture **Model** is built out of several **Element**s. Each element of a model has a given **Shape** and is needed *count* times. The table **ShapeHole** indicates the coordinates of the holes that have to be drilled in a shape (referenced by *shapeNo*). It has an attribute *seq* that indicates the order in which the holes of the given shape have to be drilled.

Questions

Provide the following query statements and commands for a PostGIS database:

1. What would be the most adequate geometric type to describe a shape? Give the command that adds that attribute to the table **Shape**. (Note that, conventionally, the SRID for a default cartesian coordinate system is -1.)
2. Give the identifier (*sNo*) of the shapes for which there exists at least one hole that is closer than 4mm from its border. (Remember that the diameter of the holes is 5mm.)
3. For a given furniture model, a good indication about how long it takes to manufacture it is given by the number of shapes it contains, the total cutting length of all these shapes and, the total number of holes that have to be drilled. Give this information for each model of the database.
4. For each shape, what are the minimal chipboard dimensions (width and length) to be used?

You might need some of the following PostGIS functions:

- `AddGeometryColumn(table_name, column_name, srid, type, dimension)`, where `type` can be any of POINT, LINE, MULTILINE, MULTIPOLYGON, etc.
- `ST_X(geom)` and `ST_Y(geometry)` return respectively the *x* and *y* coordinates of the geometry `geom`.
- `ST_PointN(geom, n)` returns the *n*-th point in a linestring or circular linestring in the geometry `geom`.
- `ST_Envelope(geom)` returns a geometry representing the bounding box of the supplied geometry: ((MINX, MINY), (MINX, MAXY), (MAXX, MAXY), (MAXX, MINY), (MINX, MINY)).
- `ST_DWithin(geom1, geom2, dist)` returns true if the geometries are within the specified distance of one another.
- `ST_Length(geom)` returns the 2d length of the geometry if it is a linestring or multilinestring.