

**CENTRALE
LYON**

PROJETS

Module ouvert disciplinaire 11.1
SYSTEMES DE GESTION DE
BASES DE DONNEES

Ecole Centrale de Lyon

Mohsen ARDABILIAN

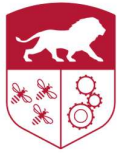
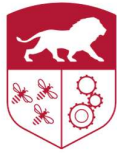


Table des matières

1. Organisation.....	1
1.1. Conception du modèle de données.....	1
1.2. Implantation et alimentation d'une base de données.....	1
1.3. Réalisation d'interfaces.....	2
1.4. Détail des livrables attendus.....	2
2. Les projets	3
Projet 1 - Gestion des rallyes sportifs.....	4
Description.....	4
Requêtes	5
Projet 2 - Library database	6
Description.....	6
Queries	7
Projet 3 - Basketball	8
Description.....	8
Queries	8
Projet 4 - Gestion des conférences	10
Description.....	10
Requêtes	11
Projet 5 - Blue Jeans Factory	13
Description.....	13
Queries	13



1. Organisation

Cinq sujets de projet sont proposés. Le travail se fait en binôme. Plusieurs binômes peuvent choisir le même sujet. Néanmoins dans ce dernier cas, chaque binôme fait le travail de bout en bout sans échanger avec les autres binômes. N.B. Il existe à ce propos des outils efficaces pour évaluer le degré de ressemblance entre plusieurs rendus...

Il est aussi possible de proposer un sujet. Dans ce cas, la complexité du modèle de données devra être du même ordre de grandeur que celle des applications proposées et le sujet devra être validé en amont par le responsable de BE. On s'attend au minimum à tous les types de relations, à la gestion de clés étrangères, et à la notion d'héritage entre certaines entités comme par exemple entre organisateurs ou concurrent et personnes, quelle que soit son implantation finale.

1.1. Conception du modèle de données

Il s'agit dans cette phase de réaliser une modélisation Entité-Association complète et correcte du modèle de données retenu. Le rapport devra comprendre :

- le diagramme Entité-Association réalisé avec DBMAIN,
- la définition textuelle des types d'entité et d'association,
- les contraintes d'intégrité.

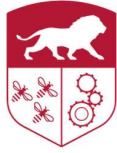
1.2. Implantation et alimentation d'une base de données

Il faut ensuite traduire le schéma conceptuel en un schéma relationnel et créer les requêtes optimisées en SQL permettant de créer les tables, les vues, les index et les fonctions SQL éventuelles.

Dans une seconde étape il s'agira d'alimenter une base avec des données en volume suffisant pour pouvoir effectuer ensuite des extractions pertinentes grâce aux requêtes, vues optimisées. Vous avez le libre choix de la base de données : SQLite, MySQL, PostgreSQL...

Le rapport fournira le schéma relationnel et les requêtes de création des entités de la base (DDL). Sans être nécessairement détaillés dans le rapport, les programmes et les requêtes nécessaires à l'alimentation de la base en données devront toutefois faire partie des livrables.

Suggestion : le code développé lors de cette phase pourrait être mis en œuvre via un ou plusieurs notebooks Jupyter.



1.3. Réalisation d'interfaces

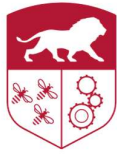
Dans cette dernière partie du travail il s'agira de réaliser une ou plusieurs interfaces applicatives permettant d'interagir avec la base de données. On peut par exemple penser à une interface d'administration, une interface utilisateur, etc.

La technologie retenue pour ces interfaces est au choix : web, Tk... L'utilisation d'un ORM et la gestion de droits d'accès en fonction du profil des utilisateurs sont un plus.

1.4. Détail des livrables attendus

Les livrables des projets devront comprendre (liste non exhaustive) :

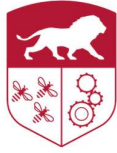
- I. un rapport comportant l'analyse et la conception, la réalisation :
 - le schéma conceptuel corrigé s'il y a lieu,
 - les tables relationnelles,
 - les vues,
 - les requêtes SQL permettant de créer les tables avec leurs contraintes,
 - les requêtes SQL d'accès aux données (CRUD) spécifiées dans chaque sujet et mises en œuvre via les interfaces de l'application
- II. les codes commentés, comprenant l'ensemble des scripts de création, d'alimentation, et de maintenance éventuelle de la base de données,
- III. la base de données, avec les données, dans un état cohérent et opérationnel, éventuellement sous la forme d'un dump SQL. Il faudra dans ce cas fournir un script et les instructions permettant l'initialisation de la base à partir du dump.
- IV. les codes de l'application permettant d'interagir avec la base,
- V. les manuels d'installation et d'usage.



**CENTRALE
LYON**

Module ouvert disciplinaire 11.1
Systèmes de gestion de bases de données

2. Les projets



Projet 1 - Gestion des rallyes sportifs

Description

Une société responsable de l'organisation de manifestations sportives veut se doter d'une base de données relationnelle contenant toutes les informations nécessaires à la gestion des rallyes de véhicules de tout type (Automobile, Camion et Moto).

Chaque rallye est identifié par son nom, son année d'organisation, ses sponsors et ses fournisseurs. Un rallye est divisé en étapes.

Chaque étape est identifiée par un numéro chronologique, la ville de départ, la ville d'arrivée, ainsi que le nombre de kilomètres à parcourir entre ces deux villes (il est à noter que les villes de départ et d'arrivée d'une étape peuvent être les mêmes).

On distingue deux types d'étapes : les liaisons et les spéciales.

Une étape liaison est chronométrée depuis la ville de départ jusqu'à la ville d'arrivée mais les concurrents n'ont pas de limite de temps à observer pour terminer l'étape.

Dans une étape spéciale, les concurrents doivent terminer dans une tranche horaire définie à l'avance par les organisateurs. Dans le cas où un concurrent termine en retard (dépassement de la tranche), il est disqualifié.

Une étape peut traverser plusieurs pays (la base de données devra permettre de retrouver facilement toutes les étapes se déroulant dans un pays donné).

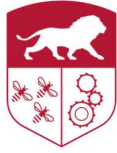
Pour chaque concurrent, on connaît son nom, son prénom, son adresse, son nombre de participations aux rallyes ainsi que sa nationalité. Chaque concurrent participe au rallye au sein d'un équipage.

Un équipage est identifié par un numéro et est composé de deux concurrents. Il est clair qu'un concurrent ne peut être inscrit que dans un seul équipage.

Chaque équipage fait partie d'une seule équipe. Les équipes sont identifiées par un nom. A chaque équipe, on associe un budget, ainsi que la liste de ses sponsors (un sponsor peut aussi bien sponsoriser des équipes autos, des équipes camions et des équipes motos).

Chaque véhicule est identifié par un numéro attribué par les organisateurs pour les différencier tout au long du rallye, une marque (TOYOTA, HONDA,...), sa cylindrée ainsi que son numéro de série. Un véhicule est affecté à un et un seul équipage (un équipage ne pouvant courir que sur un seul véhicule).

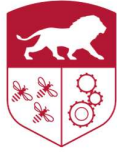
Les équipes peuvent être soit officielles (elles représentent une seule marque) ou bien une association : elles représentent dans ce cas-là un panachage de marques (HONDA_MAZDA,...).



Les résultats de chaque étape et pour chaque équipage sont classés. Un équipage pouvant être disqualifié.

Requêtes

- 1) Lister par ordre alphabétique les participants du rallye Paris Dakar de l'année 2000 ayant appartenu à la catégorie moto.
- 2) Lister le nombre d'étapes se déroulant au Sénégal.
- 3) Lister la liste d'étapes (numéro, ville départ et ville d'arrivée) se déroulant au Sénégal.
- 4) Trouver la liste des rallyes ayant un nombre de participants égal ou supérieur au nombre de participants du rallye Paris-Dakar de l'année 2002.
- 5) Lister le nombre de véhicules appartenant à la marque Toyota ayant participé au rallye Paris-Dakar pour les dix dernières années.
- 6) Lister le classement des équipages par étape pour le Paris-Dakar de l'an 2002.
- 7) Donner le nombre de participation au rallye Paris-Dakar par concurrent.



Projet 2 - Library database

Description

A single library database is shared between multiple labs of ECL. The library database stores the following information:

We distinguish different kinds of publications: regular books, periodics and internal reports:

- *Regular books* have ISBN (identifies the book), Title, Authors (assume a minimum of one and maximum of 4 authors), Publisher, Edition, Year of Publication. Books may further be classified to various categories (assume a maximum of 4 categories per book).
- *Periodics* have a volume number, a Publisher, Edition, Year of Publication.
- *Internal reports* regroup two categories of reports : ECL thesis and scientific reports. In both cases, we want to register Title, Authors (for the Phd Thesis, there is a single author), an identification number and Year of Publication.

For each book and periodic we also want to keep the book shop where it was bought and the price. Cost can be in either £, \$ or € (need to store current exchange rates, in order to do cost accounting.)

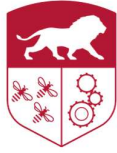
Any kind of publication has also a set of associated key words.

A particular publication is owned by a single lab. But different labs may own copies of the same publication. For simplicity, assume that only a single copy of any publication is owned by a particular lab.

The library users are identified by their Email addresses. Any user has an associated access right. That is to say, a particular user can only issue books owned by a list of Labs which authorize him/her to do so. A user has also a set of key words representing the kinds of publications he is interested in. We can imagine that he receives email when a new book corresponding to his interest is bought.

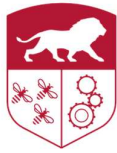
A library user may also propose new publications to be bought: he has to give all the identification information for the new publication (the same as define above for regular books, periodics and internal reports) and the date of proposal.

Publications have the following availability status: {On rack, issued to, lost, to be bought}.



Queries

- 1) List all the publications registered in the library system (Do not show duplicates owned by different labs).
- 2) For a given user, list all the publications issued to her/him and owned by any lab/ a particular lab.
- 3) Evaluate the price of all publications owned by a particular lab in €. (at the running rate of other currencies)
- 4) For a given user, evaluate whether (s)he can borrow a particular publication at a given time. (will depend on the availability status and the access rights)
- 5) If there is a publication such that a particular user has rights to borrow it (a copy), but it is (all copies for which the user has rights are) issued to some one else then show the email address(es) of all those users who presently have a borrowed copy of the publication that this user has also right to borrow.
- 6) List all publication belonging to a particular category, and costing less than a particular amount.
- 7) List all publications authored by a particular author, and published after a particular year.
- 8) For a given publisher, prepare a list of “books published” chronologically.
- 9) List all lost regular books (title, publisher, ISBN) along with owner and price, sorting them according to owner, and then ISBN.



Projet 3 - Basketball

Description

The goal of this project is to create a database to manage information about basketball players who play in the clubs or national teams from different countries.

Player: For each player we store his name, the date of birth, the height and the citizenship. Each player plays in one of the clubs. The player can change the clubs. A player can also play in the national team, but only if he is a citizen of the corresponding country.

For the players the following information is of interest: number of points for each game (3 pts, 2 pts or free throws) and the percentage of success for the points, assists, rebounds, blocks and fouls.

National Team: National teams can participate in Championships (ex: World Championship, European Championship)

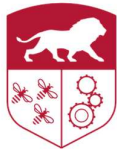
Club: For each club we store the name and the city. Clubs can belong to a league and participate in league competitions (ex: Euroleague)

Game: A game can be a part of league competition or a championship. It has two participants that are either clubs or national teams. If it is a part of league competition the season is stored.

Sponsor: The clubs and the teams are sponsored by a sponsor. They can have many sponsors and a sponsor can sponsor many clubs or teams. For each sponsor we store his name (or company name), the city and the amount. The clubs and the teams can also change the sponsors.

Queries

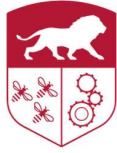
- 1) List the first 10 players who have the greatest total number of points for their national teams ever?
- 2) List the first 3 players who have the greatest free throws percentage in a final of the European Championship 2002?
- 3) Which club has the highest average height?
- 4) Which sponsor has sponsored the most number of national teams that won the World Championship?
- 5) For each club, list the players who have the greatest percentage of 3 points in the current season?



**CENTRALE
LYON**

Module ouvert disciplinaire 11.1
Systèmes de gestion de bases de données

- 6) For a particular club, which player has the most number of assists per game?
- 7) List all clubs who won the Euroleague more than 3 times?



Projet 4 - Gestion des conférences

Description

Notre laboratoire souhaite créer une base de données permettant de gérer les appels à communication pour des conférences. Nous disposons des informations suivantes :

Nous recevons plusieurs appels à communication pour des conférences chaque année. Pour chaque conférence on veut garder les informations suivantes :

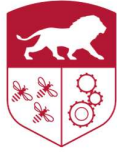
- son organisateur (généralement une université: nom, adresse, contact email),
- son intitulé,
- la date de la conférence (de telle date à telle date),
- sa localisation (ville et pays),
- la série de conférences à laquelle elle appartient (ex : SDH, CAISE, etc.),
- son texte introductif,
- la liste de responsables, avec pour chaque responsable: son nom, son prénom, son adresse professionnelle, son email et son type de responsabilité : "Program Committee chair", "Steering Committee ", "Poster chair", "General chair", "Industrial session chair",...
- l'éditeur des actes de la conférence.

Une conférence peut comprendre une ou plusieurs sessions. Pour chaque session, Nous navons :

- l'intitulé de la session,
- un ensemble de personnes qui vont lire et évaluer les soumissions (ces personnes appartiennent au comité de programme (Program committee)),
- une liste de thèmes de recherche sur lesquels porte la session. Par exemple: temporal data models, temporal database design, temporal query languages, indexing of temporal data, etc.

On veut aussi conserver les informations concernant les différentes catégories de soumission possibles pour cette conférence: "Regular paper", "panel", "tutorial", "industrial development", "system descriptions", "poster", "workshop", etc...

A chaque catégorie de soumission correspondent des spécifications différentes en terme de :



- nombre maximum de pages autorisées,
- mise en page: font, taille de caractère, type de logiciel utilisé,
- date limite de soumission,
- date de la notification de l'acceptation,
- date limite d'envoi de la version corrigée du papier.

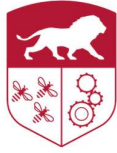
Ces spécifications peuvent être différentes suivant la catégorie de soumission : par ex, la date limite de soumission pour un tutorial et un papier n'est pas la même; pour un tutorial, un panel ou une proposition de workshop, il n'y a pas de spécification de nombre de pages.

En conjonction avec la conférence peuvent être organisés des workshops qui ont en général lieu juste avant ou après la conférence. Ces workshops seront gérés dans notre base de données comme une conférence (il y a un appel à communication : dates limites, mots clés, un comité de programme etc.). On souhaite cependant garder un lien vers la conférence à laquelle le workshop est rattaché.

Enfin, nous souhaitons gérer les utilisateurs de la base de données : Pour chaque utilisateur on conserve son nom, prénom et adresse email ainsi que son profil. Le profil d'un utilisateur correspond à un ensemble de mots clés correspondant aux thèmes qui l'intéresse. Ainsi un utilisateur pourra être informé des conférences qui correspondent à son domaine d'intérêt.

Requêtes

- 1) Afficher la liste des conférences qui ont lieu dans l'année courante.
- 2) Afficher la liste des pays, avec pour chaque pays la liste des conférences (intitulé) qui y ont été organisées.
- 3) Afficher les conférences qui portent sur un domaine de recherche portant sur les bases de données spatiales (i.e. le mot clé entré dans la base de données n'est pas forcément "bases de données spatiales").
- 4) Afficher la liste des conférences qui intéressent un utilisateur particulier.
- 5) Afficher la liste des conférences du groupe de conférence SDH.
- 6) Afficher la liste des workshops organisés conjointement avec une conférence donnée.
- 7) Afficher la liste des conférences qui ont publié dans les actes d'un éditeur donné.



**CENTRALE
LYON**

Module ouvert disciplinaire 11.1
Systèmes de gestion de bases de données

- 8) Afficher la liste des utilisateurs de la base de données qui ont aussi été responsables d'une conférence (comme responsable ou membre du comité de programme).
- 9) Afficher la liste des conférences pour lesquelles le professeur "X" a été responsable (comme responsable ou membre du comité de programme).

Projet 5 - Blue Jeans Factory

Description

Cotton Blue is a jeans factory. To manufacture a jean, employees have to do three different activities or phases of production: cut of posterior pieces, cut of frontal pieces and assembling of them. In this factory there are two types of employees. We have the following information:

Every employee is identified by a numeric code. There are employees that work "by production" and employees that work "by journey of work":

Employees that work "by production" have an associated activity and a number of pieces to produce.

The employees that work "by journey" have a associated salary.

The journey of work is of 9 hours from Monday to Friday and of 3 hours on Saturday. On Saturday the journey of work is divided by three (three phases of three hours each one).

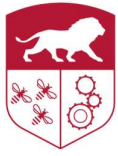
Any worker that arrives late to the factory for 50 minutes or less, will have a diminution of 5 dollars of their week salary. Workers "by journey" will have this diminution by hour and until one day if they don't come to work . Workers "by production" will have this diminution by number of pieces they didn't produced.

Workers will have a grant of the double of salary if they work more during the week. In the case of workers by production it will be according to the number of pieces produced. In the case of workers by journey it will be according to the number of extra hour worked.

For each employee we must keep his entry date in the company. Every time that a worker has one year in the company they will receive an extra salary of a week. Each worker that has been working for a year or more at the company will receive an extra salary of one month as Christmas present. Workers that have been working for less than a year at the company will receive the proportional part of this grant.

Queries

- 1) List all the workers and their corresponding salary by week.
- 2) List all the workers "by journey" that have been working since more than 3 years in the factory.
- 3) List all the workers that have had a diminution of their salary by week.
- 4) For a given worker, list all the hours that he had worked by week or the number of pieces produced according to his case.
- 5) List the total of workers that will receive the total of Christmas grant or the extra salary of a week.



- 6) List the total of workers that will receive only a proportional part of the Christmas grant.
- 7) List the total of workers that will receive an extra salary by week.