4.2.2 Relationships .................................................. 87
4.2.3 Weak entity types .............................................. 94
4.2.4 Alternative Notations ......................................... 96
4.2.5 Enhanced Entity–Relationship Model ....................... 99
4.2.6 Reverse Engineering .......................................... 99
4.3 E.R.-to-Relational Models Mapping ............................ 100
4.3.1 Intro ............................................................. 100
4.3.2 Algorithm ......................................................... 100
4.3.3 Outro ............................................................. 105
4.4 Guidelines and Normal Form ..................................... 105
4.4.1 General rules .................................................... 105
4.4.2 Example .......................................................... 106
4.4.3 Functional dependencies ..................................... 107
4.4.4 Normal forms and keys ....................................... 109
4.5 Unified Modeling Language Diagrams ........................... 112
4.5.1 Overview ......................................................... 112
4.5.2 Types of diagrams ............................................. 113
4.5.3 Zoom on classes diagrams .................................... 115
Exercises .............................................................. 117
Solution to Exercises ................................................. 122
Problems ............................................................... 127
Solution to Selected Problems ....................................... 136

5 Databases Applications .............................................. 143
Resources ............................................................... 143
5.1 Overview ............................................................. 143
5.2 Java’s Way ........................................................... 144
5.3 Flash Intro to Java .................................................. 144
5.4 A First Program ..................................................... 145
5.4.1 The Java code ................................................... 145
5.4.2 The database ..................................................... 146
5.4.3 The result ........................................................ 147
5.4.4 A variation ....................................................... 147
5.5 Mapping Datatypes .................................................. 148
5.6 Differences Between executeQuery, executeUpdate and execute ................. 148
5.7 A Second Program .................................................. 148
5.7.1 Passing options .................................................. 148
5.7.2 Creating a table .................................................. 149
5.7.3 Inserting values .................................................. 149
5.7.4 Prepared statements .......................................... 149
5.7.5 Advanced Statements objects ................................. 149
Exercises .............................................................. 149
Solution to Exercises ................................................. 151
5.8 Problems ............................................................. 152
Solution to Selected Problems ....................................... 159

6 A Bit About Security ................................................ 160
Resources ............................................................... 160
6.1 Usual Aspects ........................................................ 160
6.1.1 Threat model ..................................................... 160
6.1.2 Control measures ................................................. 160
6.2 Specificities Of Databases ........................................ 160
   6.2.1 Attack ......................................................... 160
   6.2.2 Protections .................................................... 161

7 Presentation of NoSQL .............................................. 162
   Resources .......................................................... 162
   7.1 A Bit of History .................................................. 162
      7.1.1 DB applications and application DB ...................... 162
      7.1.2 Clusters, clusters... ...................................... 162
      7.1.3 A first shift ............................................... 163
      7.1.4 Gathering forces .......................................... 163
      7.1.5 The future? ................................................ 163
   7.2 Comparison ........................................................ 164
      7.2.1 Overview ..................................................... 164
      7.2.2 ACID Vs CAP ............................................... 164
   7.3 Categories of NoSQL Systems ................................... 164
   7.4 MongoDB .......................................................... 165
      7.4.1 Resources .................................................... 165
      7.4.2 Introduction ............................................... 166
      7.4.3 Document-oriented database ................................ 166
      7.4.4 General organization of MongoDB databases .............. 167
      7.4.5 First elements of syntax .................................. 168
      7.4.6 Set up ........................................................ 169
      7.4.7 Example ..................................................... 169
   7.5 Principles .......................................................... 170
   Exercises ............................................................. 170
   Solution to Exercises .............................................. 171

References ............................................................. 172
## List of Problems

<table>
<thead>
<tr>
<th>Section</th>
<th>Problem Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>A database catalog for a campus</td>
<td>17</td>
</tr>
<tr>
<td>2.1</td>
<td>Finding candidate key in a CLASS relation</td>
<td>27</td>
</tr>
<tr>
<td>2.2</td>
<td>Relational model for a cinema company</td>
<td>28</td>
</tr>
<tr>
<td>2.3</td>
<td>Relational model for bills</td>
<td>28</td>
</tr>
<tr>
<td>3.1</td>
<td>Discovering the documentation</td>
<td>58</td>
</tr>
<tr>
<td>3.2</td>
<td>Creating and using a simple table in SQL</td>
<td>59</td>
</tr>
<tr>
<td>3.3</td>
<td>Duplicate rows in SQL</td>
<td>60</td>
</tr>
<tr>
<td>3.4</td>
<td>Revisiting the PROF table</td>
<td>60</td>
</tr>
<tr>
<td>3.5</td>
<td>TRAIN table and more advanced SQL coding</td>
<td>61</td>
</tr>
<tr>
<td>3.6</td>
<td>Insert, Select and Project in the COFFEE database</td>
<td>63</td>
</tr>
<tr>
<td>3.7</td>
<td>Select queries for the COMPUTER table</td>
<td>65</td>
</tr>
<tr>
<td>3.8</td>
<td>Improving a role-playing game</td>
<td>66</td>
</tr>
<tr>
<td>3.9</td>
<td>A simple database for books</td>
<td>67</td>
</tr>
<tr>
<td>4.1</td>
<td>Design for your professor</td>
<td>127</td>
</tr>
<tr>
<td>4.2</td>
<td>Reading the MOVIES database ER schema</td>
<td>127</td>
</tr>
<tr>
<td>4.3</td>
<td>ER diagram for car insurance</td>
<td>128</td>
</tr>
<tr>
<td>4.4</td>
<td>Reverse engineering by hand</td>
<td>129</td>
</tr>
<tr>
<td>4.5</td>
<td>Discovering MySQL Workbench</td>
<td>129</td>
</tr>
<tr>
<td>4.6</td>
<td>ER-to-Relation mapping for car insurance</td>
<td>130</td>
</tr>
<tr>
<td>4.7</td>
<td>From E.R. diagram to Relational model – BIKE</td>
<td>130</td>
</tr>
<tr>
<td>4.8</td>
<td>ER-to-Relation mapping for Country</td>
<td>131</td>
</tr>
<tr>
<td>4.9</td>
<td>From business statements to E.R. diagram – UNIVERSITY</td>
<td>131</td>
</tr>
<tr>
<td>4.10</td>
<td>Normal form of a CAR_SALE relation</td>
<td>132</td>
</tr>
<tr>
<td>4.11</td>
<td>Normalizing the FLIGHT relation</td>
<td>132</td>
</tr>
<tr>
<td>4.12</td>
<td>From business statement to dependencies</td>
<td>133</td>
</tr>
<tr>
<td>4.13</td>
<td>Normalization</td>
<td>133</td>
</tr>
<tr>
<td>4.14</td>
<td>Normal form of the BOOK relation</td>
<td>133</td>
</tr>
<tr>
<td>4.15</td>
<td>Normal form of the CONTACT relation</td>
<td>134</td>
</tr>
<tr>
<td>4.16</td>
<td>CONSULTATION relation: justification, primary key and normal form</td>
<td>134</td>
</tr>
<tr>
<td>4.17</td>
<td>From E.R. to relational schema and UML class diagram – CAR_INFO</td>
<td>134</td>
</tr>
<tr>
<td>4.18</td>
<td>Using MySQL Workbench’s reverse engineering</td>
<td>135</td>
</tr>
<tr>
<td>4.19</td>
<td>From business statements to dependencies – KEYBOARD</td>
<td>135</td>
</tr>
<tr>
<td>4.20</td>
<td>From UML to relational model – DRIVER</td>
<td>136</td>
</tr>
<tr>
<td>5.1</td>
<td>MySQL’s batch mode and HW_EBOOKSHOP.sql</td>
<td>152</td>
</tr>
<tr>
<td>5.2</td>
<td>First Database Application</td>
<td>152</td>
</tr>
<tr>
<td>5.3</td>
<td>Advanced Java Programming</td>
<td>153</td>
</tr>
</tbody>
</table>
Preamble

How to use this guide

These lecture notes are written in an elusive style: their are a support for the explanations I will be making at the board. They are not designed to be self-contained: they are rather a list of topics and reminders, along with handy examples, code and drawings. Reading them before coming to the lecture will help you getting a sense of the next topic we will be discussing, but you may sometimes have trouble deciphering their ... unique style.

When it comes to code, you can normally copy-and-paste it from the document and use it as it is. Some portion of code starts with a path in comment: you can browse the source code at https://rocketgit.com/user/caubert/CSCI_3410/source/tree/branch/master/tree/notes/code to download it directly.

On top of the notes, you will find in this document:

- References, at the very end of this document
- and for each chapter,
  - A list of additional resources,
  - A list of short exercises,
  - Solution to those exercises,
  - A list of problem,
  - Sometimes, solution to some of those problems.

Any feedback is greatly appreciated. Please refer to http://spots.augusta.edu/caubert/db/ln/README.html#contributing for how to contribute to those notes.

The syllabus is at http://spots.augusta.edu/caubert/db/, and the webpage for this notes is at http://spots.augusta.edu/caubert/db/ln/.

Planned Schedule

A typical (metting twice a week, ±17 weeks, ±30 classes) semester is divided as follows:

- Lecture 1: Presentation and Syllabus
- Lecture 2–3: Introduction
- Lecture 4–5: The Relational Model
- Lecture 6–9: The SQL Programming Language
- Lecture 10–11: Review session and Exam #1
- Lecture 12: Return on Exam + Introduction to High-Level Design
- Lecture 13–15: Entity-Relationship Model
- Lecture 16: E.R.-to-Relational Models Mapping
- Lecture 17–20: Guidelines and Normal Form
- Lecture 23–24: Review session and Exam #2
Contents

• Lecture 25: Intro to Database Programming and to Java
• Lecture 26–28: Database Applications
• Lecture 29: Introduction to NoSQL
• Lecture 30: Introduction to MongoDB

For information purposes, an indication like this:

marks the (usual) separation between two lectures.

Previous Exams

To give you a sense of what you will be asked to do during the exams, please find below a description of the exams given previous semesters.

Fall 2017

• Exam #1:¹
  – Six small exercises (Exercise 1.8, Exercise 2.4, Exercise 2.8, Exercise 3.5, Exercise 3.7 and Exercise 3.10)
  – Problem 2.1 (Finding candidate key in a CLASS relation)
  – Problem 2.2 (Relational model for a cinema company)
  – A variation on (Elmasri and Navathe 2010, Exercise 3.11), (Elmasri and Navathe 2015, Exercise 5.11)
  – Problem 3.5 (TRAIN table and more advanced SQL coding)
• Exam #2:
  – Six small exercises,
  – Problem 4.9 (From business statements to E.R. diagram – UNIVERSITY)
  – Problem 4.12 (From business statement to dependencies)
  – Problem 4.15 (Normal form of the CONTACT relation)
  – A variation on Problem 4.8 (ER-to-Relation mapping for Country)
• Final:
  – A variation on (Exercise 5.10)
  – A variation on Problem 3.6 (Insert, Select and Project in the COFFEE database)
  – A variation on Problem 4.20 (From UML to relational model – DRIVER): students were asked to draw the ER diagram for that schema.
  – Problem 4.14 (Normal form of the BOOK relation)
  – Problem 4.16 (CONSULTATION relation: justification, primary key and normal form)

Spring 2018

• Exam #1:
  – Problem 2.2 (Relational model for a cinema company), except that I gave some of the relations and attributes, to help getting started with the problem.
  – Problem 4.1 (Design for your professor)
  – Problem 3.5 (TRAIN table and more advanced SQL coding)
• Exam #2:

¹This exam was probably a bit too long, but students managed it pretty well.
A variation on Problem 4.15 (Normal form of the CONTACT relation)
A variation on Problem 4.16 (CONSULTATION relation: justification, primary key and normal form)
Problem 3.8 (Improving a role-playing game)
A variation on Problem 4.19 (From business statements to dependencies – KEYBOARD)
Problem 4.8 (ER-to-Relation mapping for Country)
• Final:
  – Take the relational model of the solution of Problem 2.2 (Relational model for a cinema company), and "reverse-engineer" it to obtain a ER diagram (this problem was harder than I expected)
  – Six small exercises (Exercise 4.24, Exercise 7.1, Exercise 7.2, Exercise 7.3, Exercise 7.4)
  – Problem 4.11 (Normalizing the FLIGHT relation)
  – A variation on Problem 4.9 (From business statements to E.R. diagram – UNIVERSITY)
  – A variation on Problem 4.20 (From UML to relational model – DRIVER): students were asked to draw the ER diagram for that schema.

Fall 2019

– Exam #1:
  – Problem 3.9 (A simple database for books)
  – Five exercises (Exercise 1.6, Exercise 2.12, Exercise 3.29, Exercise 2.13, Exercise 3.36)
  – A variation on Problem 2.3 (Relational model for bills)

Typesetting and Acknowledgments

The source code for those notes is hosted at rocketgit², typeset in markdown, and then compiled using pandoc³ and multiple filters (pandoc-numbering⁴, pandoc-citeproc⁵, pandoc-include-code⁶). The document uses Linux Libertine fonts⁷, and the drawings use various LaTeX packages, including PGF, TikZ⁹, tikz-er²¹ and tikz-dependency²². The help from the TeX - LaTeX Stack Exchange¹³ community greatly improved this document. The underline text is obtained from https://yaytext.com/underline/, the unicode symbols are searched in http://milde.users.sourceforge.net/LUCR/Math/unimathsymbols.html.

To clone and compile this document, refer to the README.md file, at https://rocketgit.com/user/caubert/CSCI_3410/source/tree/branch/master/blob/README.md or http://spots.augusta.edu/caubert/db/ln/README.html.

²https://rocketgit.com/user/caubert/CSCI_3410
³http://pandoc.org/
⁴https://github.com/chdemko/pandoc-numbering
⁵https://github.com/jgm/pandoc-citeproc
⁶https://github.com/owickstrom/pandoc-include-code
⁷http://libertine-fonts.org/
⁸At least, the pdf version.
⁹https://www.latex-project.org/
¹⁰https://sourceforge.net/projects/pgf/
¹¹https://bitbucket.org/pavel_calado/tikz-er2/raw/da9f9f7f169647cad6d91df7975400b1605ae67a/tikz-er2.sty
¹²https://ctan.org/pkg/tikz-dependency
¹³https://tex.stackexchange.com/
Those lecture notes were created under an Affordable Learning Georgia\textsuperscript{14} Mini-Grant for Ancillary Materials Creation and Revision\textsuperscript{15} (Proposal M71\textsuperscript{16}).

\section*{Various Resources}

On top of the references and of the “resources” listed at the beginning of each chapter, I consulted while writing those notes:

- https://www.1keydata.com/datawarehousing/data-modeling-levels.html
- https://creately.com/blog/diagrams/class-diagram-relationships/
- https://www.ntu.edu.sg/home/ehchua/programming/java/JDBC_Basic.html

\section*{Copyright}

This work is under Creative Commons Attribution 4.0 International License\textsuperscript{17} or later.

Some figures and resources are borrowed from other sources, like (Elmasri and Navathe 2010), (Elmasri and Navathe 2015), or the resources mentioned previously, in which case it is indicated clearly.

\section*{Softwares}

*I would like here to include some of the softwares I believe could make your semester easier, but still have to work on it.*
1 Introduction

Resources

- (Elmasri and Navathe 2010, ch. 1.1–1.6)
- (Elmasri and Navathe 2015, ch. 1.1–1.6)

1.1 Database

A database (DB) is a collection of related data. Data (= information, can be anything, really) + management (= logical organization of the data), through Database Management System).

1. Represent a mini-world, a Universe of Disclosure (UoD).
2. Logically coherent, with a meaning.
3. Populated for a purpose.

Refer to the “A simplified database environment” figure, where

- The program can be written in any language, be a web interface, etc.
- Most DBMS software include a Command-Line Interface (CLI),
- Sometimes, meta-data and data are closer than pictured (you can have “self-describing meta-data”, that is, they cannot be distinguished). Note also that “meta-data” has numerous definition (“data about the data”): we use it here to refer to the description of the organization of the data, and not e.g. statistical data about our data.

1.2 Database Management System (DBMS)

A DBMS is a general purpose software that is used to

1. Define (= datatype, constraints, structures, etc.)
2. Construct / Create (= storing the data)
3. Manipulate / Maintain (= query, update, etc.)
4. Share / Control access (= among users, softwares.)

You can think of a tool to

1. specify a storage unit,
2. fill it,
3. allow to change its content, as well as its organization,
4. allow multiple users to access it at the same time.

1.3 Subtasks

1. Organization (DB designer)
Figure 1.1: A simplified database environment
2. Modification, retrieval (end-user)
3. Maintenance (DB administrator)
4. Software engineer, web developer, programers, ...

The focus will be on design, but we will have to do a little bit of everything.

1.4 Design

Refer to the “The cycle of design” figure.

Add a line from SQL to Program, a line from NoSQL to CLI. Make the arrow from Bus. Statements to NoSQL dashed?

1.5 An Example

<table>
<thead>
<tr>
<th>STUDENT</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Student_number</td>
<td>Class</td>
<td>Major</td>
<td></td>
</tr>
<tr>
<td>Morgan</td>
<td>18</td>
<td>2</td>
<td>IT</td>
<td></td>
</tr>
<tr>
<td>Bob</td>
<td>17</td>
<td>1</td>
<td>CS</td>
<td></td>
</tr>
</tbody>
</table>

COURSE
### Course_list

<table>
<thead>
<tr>
<th>Course_name</th>
<th>Course_number</th>
<th>Credit_hours</th>
<th>Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intro. to CS</td>
<td>1301</td>
<td>4</td>
<td>CS</td>
</tr>
<tr>
<td>DB Systems</td>
<td>3401</td>
<td>3</td>
<td>CS</td>
</tr>
</tbody>
</table>

### SECTION

<table>
<thead>
<tr>
<th>Section_identifier</th>
<th>Course_num</th>
<th>Semster</th>
<th>Year</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2910</td>
<td>1301</td>
<td>Fall</td>
<td>2019</td>
<td>Kate</td>
</tr>
<tr>
<td>9230</td>
<td>2103</td>
<td>Spring</td>
<td>2020</td>
<td>Todd</td>
</tr>
</tbody>
</table>

### GRADE_REPORT

<table>
<thead>
<tr>
<th>Student_number</th>
<th>Section_identifier</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>2910</td>
<td>A</td>
</tr>
<tr>
<td>18</td>
<td>2910</td>
<td>B</td>
</tr>
</tbody>
</table>

### PREREQUISITE

<table>
<thead>
<tr>
<th>Course_number</th>
<th>Prerequisite_number</th>
</tr>
</thead>
<tbody>
<tr>
<td>2910</td>
<td>1301</td>
</tr>
<tr>
<td>1302</td>
<td>1301</td>
</tr>
</tbody>
</table>

### 1.5.1 Structure

- Database structure and records, 5 files (=collection of records), each containing data records of the same type. **Persistent storage**
- Each record has a structure, different data elements, each has a data type.
- Records have relationships between them.

### 1.5.2 Interactions

- Can I retrieve the name of 1301? Can I know what classes Kate is teaching this semester? Can I know what instructor Bob had?
- Queries, updates, removal, addition of records. **Efficiency** (using auxiliary files (indexes), optimization)
- Selection (for any operation) requires care: do we want all the records, some of them, exactly one?

### 1.5.3 Organization

Why are the files separated like that? Why don’t we store the section with the course with the students?

- **Avoiding redundancy** (“data normalization”), or having it controlled
- **Levels of access** (multiple user interface)
• And we still have the same usability!

But need to be careful about **consistency / referential integrity**.

### 1.5.4 How is a database conceived?

- Specification and analysis. “Each student number will be unique, but they can have the same name. We want to access the letter grade, but not the numerical grade”, etc.
- Conceptual design
- Logical design
- Physical design

Gradation, from really abstract specification that is easy to modify, to more solidified description of what needs to be coded. We’ll see when we’ll study high-level models what that means. It’s easier to move things around early in the conception.

### 1.6 Characteristics of the Database Approach

1. A database is more than just data: it also contains a complete description of the structure and constraints. ⇒ We have Data & Meta-data (we can also have self-describing data, where meta-data and data are interleaved).
2. Data-abstraction: DBMS provides a conceptual representation, and hides implementation details.
   a) Program-data independence: changing the database doesn’t require to change the DBMS. Compare with changing a custom data-type in a program.
   b) Program-operation independence: an operation has an interface (or signature) and an implementation (or method)
3. Support of multiple views of the data: view is a subset of the database, or virtual data.
4. Sharing and multiuser transaction processing: concurrency control using transactions (= series of instructions that is supposed to execute a logically correct database access if executed in its entirety). Isolation, atomicity (all or nothing): ACID principles, that we will study later on.

### Exercises

**Exercise 1.1** Is a pile of trash a database? Why, or why not?

**Exercise 1.2** Which one comes first, the physical design, the conceptual design, or the logical design?

**Exercise 1.3** What is the difference between a database and the meta-data of the database?

**Exercise 1.4** Do I have to change my DBMS if I want to change the structure of my data?

**Exercise 1.5** What is a virtual data? How can I access it?

**Exercise 1.6** Name two DBMS.

**Exercise 1.7** Why do DBMS include concurrency control?

**Exercise 1.8** What is independence between program and data? Why does it matter?

**Exercise 1.9** Define the word “miniworld”.

Exercise 1.10  Expand the acronym “DBMS”.

Exercise 1.11  Name the four different kinds of action that can be performed on data.

Exercise 1.12  Assume that I have a file where one record corresponds to one student. Should the information about the classes a student is taking (e.g. room, instructor, code, etc.) being stored in the same file? Why, or why not?

Solution to Exercises

Solution 1.1  No, because it lacks a logical structure.

Solution 1.2  The conceptual design.

Solution 1.3  The data is the information we want to store, the meta-data is its organization, how we are going to store it. Meta-data is information about the data, but of no use on its own.

Solution 1.4  Normally no, data and programs are independent. But actually, this is true only if the model doesn’t change: shifting to a "less structured model", e.g., one of the NoSQL models, can require to change the DBMS.

Solution 1.5  It is a set of information that is derived from the database but not directly stored in it. It is accessed through queries.

Solution 1.6  Oracle RDBMS, IBM DB2, Microsoft SQL Server, MySQL, PostgreSQL, Microsoft Access, etc., are valid answers. Are not valid “SQL”, “NoSQL”, “Relational Model”, or such: we are asking for the names of actual softwares!

Solution 1.7  To ensure that several users trying to update the same data will do so in a controlled manner. To avoid inconsistency.

Solution 1.8  The application should not be sensible to the “internals” of the definition and organization of the data. It matters because having this independance means that changing the data will not require to change the programs.

Solution 1.9  The mini-world is the part of the universe we want to represent in the database. It is supposed to be meaningful and will serve a purpose.

Solution 1.10  Database Management System

Solution 1.11  The four actions are:

- Add / Insert
- Update / Modify
- Search / Query
- Delete / Remove

Solution 1.12  If we were to store all the information about the classes in the student records, then we would have to store it as many time as its number of students! It is better to store it in a different file, and then to "link" the two files, to avoid redundancy.
1.6 Characteristics of the Database Approach

Problems

Problem 1.1 *(A database catalog for a campus)* We want to define a CAMPUS database organized into three files as follows:

- A BUILDING file storing the name and GPS coordinates of each building.
- A ROOM file storing the building, number and floor of each room.
- A PROF file storing the name, phone number, email and room number where the office is located for each professor.

Pb 1.1 – Question 1 A database catalog is made of two part: a table containing the relations’ name and their number of columns, and a table containing the columns’ name, their data type, and the relation to which they belong. (You can see an example in (Elmasri and Navathe 2010, Figure 1.3), or (Elmasri and Navathe 2015, Figure 1.3)). Write the database catalog corresponding to the CAMPUS database.

Pb 1.1 – Question 2 Invent data for such a database, with two buildings, three rooms and two professors.

Pb 1.1 – Question 3 Answer the following, assuming all the knowledge you have of the situation comes from the CAMPUS database, which is an up-to-date and accurate representation of its miniworld:

1. Is it possible to list all the professors?
2. Is it possible to tell in which department is a professor?
3. Is it possible to get the office hours of a professor?
4. Is it possible to list all the professors whose offices are in the same building?
5. Is it possible to list all the rooms?
6. If a new professor arrives, and has to share his office with another professor, do you have to revise your database catalog?
7. Can you list which professors are at the same floor?
8. Can you tell which professor has the highest evaluations?

Solution to Selected Problems

Solution to Pb 1.1 Pb 1.1 – Solution to Q. 1 The database catalog should be similar to the following:

**RELATIONS**

<table>
<thead>
<tr>
<th>Relation_name</th>
<th>No_of_columns</th>
</tr>
</thead>
<tbody>
<tr>
<td>BUILDING</td>
<td>3</td>
</tr>
<tr>
<td>ROOM</td>
<td>3</td>
</tr>
<tr>
<td>PROF</td>
<td>4</td>
</tr>
</tbody>
</table>

**COLUMNS**

<table>
<thead>
<tr>
<th>Column_name</th>
<th>Data_type</th>
<th>Belongs_to_relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building_Name</td>
<td>Character(30)</td>
<td>Building</td>
</tr>
<tr>
<td>GPSLat</td>
<td>Decimal(9,6)</td>
<td>Building</td>
</tr>
<tr>
<td>GPSLon</td>
<td>Decimal(9,6)</td>
<td>Building</td>
</tr>
</tbody>
</table>
### 1.6 Characteristics of the Database Approach

<table>
<thead>
<tr>
<th>Column_name</th>
<th>Data_type</th>
<th>Belongs_to_relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building_Name</td>
<td>Character(30)</td>
<td>ROOM</td>
</tr>
<tr>
<td>Room_Number</td>
<td>Integer(1)</td>
<td>ROOM</td>
</tr>
<tr>
<td>Floor</td>
<td>Integer(1)</td>
<td>ROOM</td>
</tr>
<tr>
<td>Prof_Name</td>
<td>Character(30)</td>
<td>PROF</td>
</tr>
<tr>
<td>Phone</td>
<td>Integer(10)</td>
<td>PROF</td>
</tr>
<tr>
<td>Email</td>
<td>Character(30)</td>
<td>PROF</td>
</tr>
<tr>
<td>Room_Number</td>
<td>Integer(1)</td>
<td>PROF</td>
</tr>
</tbody>
</table>

#### Pb 1.1 – Solution to Q. 2
For the data, you could have:

- For the BUILDING file, we could have:
  
  (Allgood Hall, 33.47520, -82.02503),
  (Institut Galilé, 48.959001, 2.339999)

- For the ROOM file, we could have:
  
  (Allgood Hall, 128, 1),
  (Institut Galilé, 205, 3),
  (Allgood Hall, 228, 2)

- For the PROF file, we could have:
  
  (Aubert, 839401, dae@ipn.net, 128),
  (Mazza, 938130, Dm@fai.net, 205)

#### Pb 1.1 – Solution to Q. 3
If everything we knew about the campus came from that database, then

1. Yes, we could list all the professors.
2. No, we could not tell in which department is a professor.
3. No, we could not get the office hours of a professor.
4. Yes, we could list all the professors whose offices are in the same building.
5. Yes, we could list all the rooms.
6. If a new professor arrives, and has to share his office with another professor, we would not have to revise our database catalog (it is fine for two professor to have the same room number, in our model).
7. Yes, we could list which professors are at the same floor.
8. No, we could not tell which professor has the highest evaluations.
2 The Relational Model

Resources

- (Elmasri and Navathe 2010, Ch. 3), (Elmasri and Navathe 2015, Ch. 5), including the exercises (look at the exercises 15 and 16, for instance).
- The wikipedia page for Relational model\(^1\) and the category “Relational database management systems”\(^2\).

2.1 Concepts

The relational data model (or relational database schema) is:

- a mathematical model (use mathematical relations, set-theory, first-order predicate logic)
- with multiple implementations ("engineering approximation")

2.1.1 Domains, Attributes, Tuples and Relations

- **Domain** (or type) = set of atomic (as far as the relation is concerned) values. You can compare it to datatype and literals, and indeed it can be given in the form of a data type, but it can be named and carry a logical definition (i.e., `List_of_major` as an enumerated data type, instead of just `String`).
- **Attribute** = Attribute name + attribute domain (but we’ll just write the name).
- **Relation Schema** (or scheme) = description of a relation, often written "RELATION_NAME(\(\text{Attribute}_1, \ldots, \text{Attribute}_n\))", where \(n\) is the degree (arity) of the relation, and the domain of \(\text{Attribute}_i\) is written `dom(\(\text{Attribute}_i\))`.

\(^1\)https://en.wikipedia.org/wiki/Relational_model
\(^2\)https://en.wikipedia.org/wiki/Category:Relational_database_management_systems

![Figure 2.1: Terminology](image)
• **Tuple** \( t \) of the schema \( R(A_1, ..., A_n) \) is an ordered list of values \( <v_1, ..., v_n> \) where \( v_i \) is in \( \text{dom}(A_i) \) or a special NULL value.

• **Relation** (or relation state) \( r \) of the schema \( R(A_1, ..., A_n) \), also written \( r(R) \), is the set of \( n \)-tuples \( t_1, ..., t_m \) where each \( t_i \) is a tuple of the schema \( R(A_1, ..., A_n) \).

### 2.1.2 Characteristics of Relations

- In a relation, the order of tuples does not matter (a relation is a set). Order in tuple do matter (alternate representation where this isn’t true exist, cf. self-describing data).
- Value is atomic = “flat relational model”, we will always be in the first normal form (not composite, not multi-valued).
- NULL is N/A, unknown, unavailable (or withheld).
- Relation Schema = assertion (“Every student has a name, a SSN, …”). Tuple = fact (“The student Bob Taylor has SSN 12898, …”).
- Relations represents uniformly entities (STUDENT(...)) and relations (PREREQUISITE(Course_number, Prerequisite_number)).

### 2.1.3 Notation

- **STUDENT** = relation schema + current relation state
- **STUDENT(Name, ..., Major) =** relation schema only
- **STUDENT.Name =** Attribute Name in the relation **STUDENT**
- **t[Name], t[Name, Major], t.Name** (overloading the previous notation)

### 2.2 Constraints

#### 2.2.1 Types of constraints

We now study constraints on the tuples. There are constraints on the scheme, for instance, “a relation cannot have two attributes with the same name”

#### 2.2.1.1 Inherent model-based constraints (implicit)

Those are part of the definition of the relational model.

- No duplicate tuple
- Arity must match

#### 2.2.1.2 Schema-based constraints (explicit)

Those are parts of the schema.

- Value must match domain (“Domain constraint”), knowing that a domain can have constraints (NOT NULL, UNIQUE)
- Entity integrity constraint (no primary key value can be NULL)
- Referential integrity constraint (referred values must exist)

The last two will be studied in the next section.
2.2 Constraints

2.2.1.3 Application-based constraints (semantics)

Constraints that cannot be expressed in the schema, and hence must be enforced by some other way. Example: "the date of birth of an employee must be greater than xxx", "this year’s salary increase must be more than last year’s".

2.2.2 Keys

Tuples cannot be equal, so a subset of values must distinguish them, we study the corresponding subset of attributes.

- A **superkey** is the subset of attributes \( SK \) is a superkey for the relation \( R \), if for all relation state \( r \) of \( R \), all tuples \( t_1, t_2 \) in \( r \) are such that \( t_1[SK] \neq t_2[SK] \).
- A **key** is a minimal superkey (i.e., removing any attribute from \( SK \) would break the uniqueness property).
- A **candidate key** is a key, a **primary key** is the selected candidate key (it is underlined\(^3\)).

Note: here we “retro-fit” those definitions, in DB design, they come first! We are making the assumption that the data pre-exist to the specification to make the concept clearer.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yellow</td>
<td>Rectangle</td>
<td>10</td>
<td>(5, 3)</td>
<td></td>
</tr>
<tr>
<td>Blue</td>
<td>Rectangle</td>
<td>10</td>
<td>(3, 9)</td>
<td></td>
</tr>
<tr>
<td>Blue</td>
<td>Circle</td>
<td>9</td>
<td>(4, 6)</td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{array}{c|c|c|c|c}
\{A, B, C, D\} & \{A\} & \{B, C\} & \{D\} \\
\hline
\text{Superkey?} & √ & x & √ & √ \\
\text{Key?} & x & x & √ & √ \\
\end{array}
\]

2.2.3 Foreign Keys

A **foreign key** (FK) is a set of attributes denoted FK in the relation schema \( R_1 \), such that FK is a foreign key of \( R_1 \) ("referencing relation") that references \( R_2 \) ("referenced relation") if

- FK refers to \( R_2 \) (i.e., the attributes in FK have the same domain(s) as the primary key PK of \( R_2 \))
- a value of FK in a tuple \( t_1 \) of \( r_1(R_1) \) either
  - occurs as a value of PK for some tuple \( t_2 \) of \( r_2(R_2) \), i.e., \( t_1[FK] = t_2[PK] \)
  - is NULL
  - in the first case, we say that "\( t_1 \) refers to \( t_2 \)."

There is a referential integrity constraint from \( R_1 \) to \( R_2 \). We draw it with an arrow, note that it is possible that \( R_1 = R_2 \).

---

\(^3\)For technical reasons, underlined words cannot be searched in the document.
2.3 Transactions and Operations

Operations are of two kinds: retrievals and updates.

- Retrievals leave the relation state as it is and output a result relation. That is, retrieval: relation state $\rightarrow$ result relation
- Updates change the relation state. That is, update: relation state $\rightarrow$ relation state

They are two constraints for updates:

1. The new relation state must be “valid” (i.e., comply with the state constraints).
2. There might be transition constraints (your balance cannot become negative, for instance).

A transaction is a series of retrievals and updates performed by an application program, that leaves the DB in a consistent state.

2.3.1 Insert / Delete / Update

(1.), (2.) and (3.) refers to the “remedies”, discussed in the next section.

(Yes, we do need the state and the licence number to uniquely identify a driver’s licence, since many states use the same licence format. Entity Integrity Constraint and Referential Integrity Constraint.

4https://ntsi.com/drivers-license-format/
2.3 Transactions and Operations

2.3.1 Insert

Insert <109920, Honda, Accord, 2012> into CAR.

How things can go wrong:
- NULL for the primary key (1.)
- Duplicate value for the primary key (1.)
- Wrong number of arguments (1.)
- Fail to reference an existing value for a foreign key (1.)

2.3.1.2 Delete

Delete the DRIVER tuple with State = GA and Licence_number = 123

How things can go wrong:
- Deleting tuples inadvertently (meta)
- Deleting tuples that are referenced (1., 2., 3.)

2.3.1.3 Update (a.k.a. modify)

Update Name of tuple in DRIVER where State = GA and Licence_number = 123 to Georges

How things can go wrong:
- NULL for the primary key (1.)
- Duplicate value for the primary key (1.)
- Change value that are referenced (1., 2., 3.)
- Change foreign key to a non-existing value (1.)

2.3.2 Dealing with violations

1. Reject (restrict)
2. Cascade (propagate)
3. Set default, or set NULL

Go back to the CAR example and populate it with some data to “see” how those options affects how the database react to various operations.

Exercises

Exercise 2.1 Connect the dots:

Row • • Attribute
Column header • • Tuple
Table • • Relation

Exercise 2.2 What do we call the number of attributes in a relation?

Exercise 2.3 At the logical level, does the order of the tuples in a relation matter?
**Exercise 2.4** What is the difference between a database schema and a database state?

**Exercise 2.5** What should we put in an attribute if its value is unknown?

**Exercise 2.6** What, if any, is the difference between a superkey, a key, and a primary key?

**Exercise 2.7** Name the two kind of integrities that must be respected by the tuples in a relation.

**Exercise 2.8** What is entity integrity? Why is it useful?

**Exercise 2.9** Are we violating an integrity constraint if we try to set the value of a primary key to NULL? If yes, which one?

**Exercise 2.10** If in a relation $R_1$, an attribute $A_1$ is a foreign key referencing an attribute $A_2$ in a relation $R_2$, what does this implies about $A_2$?

**Exercise 2.11** Give three examples of operations.

**Exercise 2.12** What is the difference between an operation and a transaction?

**Exercise 2.13** Give three different ways to deal with operations whose execution in isolation would result in the violation of one of the constraint.

**Exercise 2.14** Consider the following two relations:

- COMPUTER(Owner, RAM, Year, Brand)
- OS(Name, Version, Architecture)

For each, give a) The arity of the relation, b) A (preferably plausible) example of tuple to insert.

**Exercise 2.15** Define what is the domain constraint.

**Exercise 2.16** Consider the following three relations:

- AUTHOR
  - Ref
  - Name
  - Address

- BOOK
  - ISSN
  - AuthorRef
  - Title

- GAINED-AWARD
  - Ref
  - Name
  - BookISSN
  - Year

For each relation, answer the following:

1. What is, presumably, the primary key?
2. Are they, presumably, any foreign key?
3. Using the model you defined, could we determine which author won the greatest number of awards a particular year?
Exercise 2.17  Consider the following three relations

1. What are the foreign keys in the ASSIGNED-TO relation? What are they referring to?

2. In the ASSIGNED-TO relation, explain why the Date attribute is part of the primary key. What would happen if it was not?

3. Assuming the database is empty, are the following instructions valid? If not, what integrity constraint are they violating?
   a) Insert <'AM-356', 'Surfliner', 2012> into TRAIN
   b) Insert <NULL, 'Graham Palmer', 'Senior'> into CONDUCTOR
   c) Insert <'XB-124', 'GPalmer', '02/04/2018'> into ASSIGNED-TO
   d) Insert <'BTed', 'Bobby Ted', 'Senior'> and <'BTed', 'Bobby Ted Jr.', 'Junior'> into CONDUCTOR

Exercise 2.18  Consider the following relation schema and state:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Blue</td>
<td>Austin</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Yellow</td>
<td>Paris</td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Purple</td>
<td>Pisa</td>
<td>false</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Yellow</td>
<td>Augusta</td>
<td>true</td>
<td></td>
</tr>
</tbody>
</table>

Assuming that this is all the data we will ever have, discuss whenever \( \{A, B, C, D\} \), \( \{A, B\} \) and \( \{B\} \) are superkeys and/or keys.

Solution to Exercises

Solution 2.1  Row is Tuple, Column header is Attribute, Table is Relation.

Solution 2.2  The degree, or arity, of the relation.

Solution 2.3  No, it is a set.
Solution 2.4 The schema is the organization of the database (the meta-data), while the state is the state is the content of the database (the data).

Solution 2.5 NULL

Solution 2.6 A superkey is a subset of attributes such that no two tuples have the same combination of values for all those attributes. A key is a minimal superkey, i.e., a superkey from which we cannot remove any attribute without losing the uniqueness constraint. The primary key is one of the candidate key, i.e., the key that was chosen.

Solution 2.7 Referential integrity and entity integrity.

Solution 2.8 Entity integrity ensures that each row of a table has a unique and non-null primary key value. It allows to make sure that every tuple is different from the others, and helps to "pick" elements in the database.

Solution 2.9 Yes, the entity integrity constraint.

Solution 2.10 Then we know that $A_2$ is the primary key of $R_2$.

Solution 2.11 Reading from the database, performing UPDATE or DELETE operations.

Solution 2.12 a) COMPUTER has for arity 4, and OS has for arity 3.

Solution 2.13 An operation is an “atomic action” that can be performed on the database (adding an element, updating a value, removing an element, etc.). A transaction is a series of such operations, and the assumption is that, even if it can be made of operations that, taken individually, could violate a constraint, an operation will leave the database in a consistent state.

Solution 2.14 An operation whose execution in isolation would result in the violation of a constraint can either a) be “restricted” (i.e., not executed), b) result in a propagation (i.e., the tuples that would violate a constraint are updated or deleted accordingly), or c) result in some values in tuples that would violate a constraint to be set to a default value, or the NULL value (this last option works only if the constraint violated is the referential entity constraint).

Solution 2.15 The requirement that each tuple must have for an attribute $A$ an atomic value from the domain $\text{dom}(A)$, or NULL.

Solution 2.16 To answer 1 and 2, the diagram would become:
For the last question, the answer is yes: based on the ISSN of the book, we can retrieve the author of the book. Hence, knowing which book was awarded which year, by looking in the GAINED-AWARD table, gives us the answer to that question.

Solution 2.17  1. In ASSIGNED-TO, TrainRef is a FK to TRAIN.Ref, and ConductorID is a FK to CONDUCTOR.CompanyID.

2. In this model, a conductor can be assigned to different trains on different days. If Date was not part of the PK of ASSIGNED-TO, then a conductor could be assigned to only one train.

3. a) Yes, this instruction is valid.
   b) No, it violates the entity integrity constraint: NULL can be given as a value to an attribute that is part of the PK.
   c) No, it violates the referential integrity constraint: 'XB-124' and 'GPalmer' are not values in TRAIN.Ref and CONDUCTOR.CompanyID.
   d) No, it violates the key constraint: two tuples cannot have the same value for the primary key.

Solution 2.18  • \{A, B, C, D\} is a superkey (the set of all the attributes is always a superkey), but not a superkey, as removing e.g. D would still make it a superkey.
   • \{A, B\} is a superkey and a key, as neither \{A\} nor \{B\} are keys.
   • \{A\} is not a key, and not a superkey: multiple tuples have the value 1.

Problems

Problem 2.1 (Finding candidate key in a CLASS relation) Consider the relation

CLASS(Course_Number, Univ_Section_Number, Instructor_Name, Semester, Building_Code, Room_Number, Time, Weekdays, Credit_Hours)

This relation represents classes taught in a university. The goal is to be able to have multiple offerings (classes) of courses over several semesters. List three possible candidate keys and
describe under what conditions each candidate key would be valid. Each candidate key should have between one and three attributes.

Here are some examples of values for the attributes:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Possible Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course_Number</td>
<td>CSCI3410, CSCI1302</td>
</tr>
<tr>
<td>Building_Code</td>
<td>AH, UH, ECC</td>
</tr>
<tr>
<td>Univ_SECTION_NUMBER</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>Room_NUMBER</td>
<td>E127, N118</td>
</tr>
<tr>
<td>Instructor_Name</td>
<td>John Smith, Sophie Adams</td>
</tr>
<tr>
<td>Time</td>
<td>1400, 1230, 0900</td>
</tr>
<tr>
<td>Semester</td>
<td>Spring 2015, Fall 2010, Summer 2012</td>
</tr>
<tr>
<td>Weekdays</td>
<td>M, MW, MWF, T, TH</td>
</tr>
<tr>
<td>Credit_Hours</td>
<td>1, 2, 3, 4</td>
</tr>
</tbody>
</table>

Problem 2.2 (**Relational model for a cinema company**) A cinema company wants you to design a relational model for the following set-up:

- The company has movie stars. Each star has a name, birth date, and unique id.
- The company has information about movies: title, year, length, and genre. Each movie has a unique id, and features multiple stars.
- The company owns movie theaters as well. Each theater has a name, address, and a unique id.
- Furthermore, each theater has a set of auditoriums. Each auditorium has a unique number, and seating capacity.
- Each theater can schedule movies at show-times. Each show-time has a unique id, a start time, and is for a specific movie, at a theater auditorium.
- The company sells tickets for scheduled show-times. Each ticket has a unique ticket id, and a price.

Problem 2.3 (**Relational model for bills**) Propose a relational model for the following situation:

- You want to store the bills that are debated and voted on at the U.S. House of Representatives. Each bill has a name, a (unique) sponsor which needs to be a member of the house, a date when it was discussed (which may not have been fixed yet).
- You want to record the name, political group, beginning and (expected) end of the term of the members of the house.
- You also want to record the name of the Speaker, Majority Leader, Minority Leader, Majority Whip, and Minority Whip, which are all members of the house.
- Finally, for each bill, you want to record the vote of every member of the house.
Solution to Selected Problems

Solution to Pb 2.2

Solution to Pb 2.3  Be careful: saying that a bill has a unique sponsor does not imply that a the sponsor is a good primary key for the bills: a house member could very well be the sponsor of multiple bills! It just implies that a single attribute is enough to hold the name of the sponsor.

For simplicity, we added an Id to our MEMBER and BILL relations. Note that having a “role” in the MEMBER relation to store the information about speaker, etc., would be extremely inefficient, since we would add an attribute to the ~435 members that would be NULL in ~430 of them.
3 The SQL Programming Language

Resources

- (Elmasri and Navathe 2010, Ch. 4–5), (Elmasri and Navathe 2015, Ch. 6–7) describes SQL, not one of its implementation.

This chapter will be "code-driven": the code will illustrate and help you understand some concepts. You may want to have a look at the “Setting Up Your Work Environment” Section, as early as possible in this lecture. On top of being a step-by-step guide to install and configure a relational database management system, it contains a list of useful links.

3.1 Actors

3.1.1 Technologies

- There are other models: document-based, graph, column-based, and key-value models. “NoSQL” data-model, more flexible, but only defined by opposition.
- Most commons DBMS are relational database management system (RDBMS):
  - Oracle Database
  - MySQL and its fork, MariaDB
  - Microsoft SQL Server
  - PostgreSQL
  - IBM DB2
  - Microsoft Access
  - SQLite

  Most of them supports semi-structured data, i.e., models that are not strictly speaking relational, some are "multi-model DBMS".
- Structured Query Language is the language for RDBMS, it is made of 4 sublanguages:

2https://www.mysql.com/
3https://mariadb.org/
5https://www.postgresql.org/
7https://products.office.com/en-us/access
8https://www.sqlite.org/index.html
3.1 Actors

- Data Query Language,
- Data Definition Language (schema creation and modification),
- Data Control Language (authorizations, users),
- Data Manipulation Language (insert, update and delete).

The three last sublanguages being dubbed “Data Manipulation Language”.

3.1.2 SQL

3.1.2.1 Yet another vocabulary

<table>
<thead>
<tr>
<th>“Common” / Relational</th>
<th>SQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Database”</td>
<td>Schema</td>
</tr>
<tr>
<td>“Set of databases”</td>
<td>Catalog (named collection of schema)</td>
</tr>
<tr>
<td>Relation</td>
<td>Table</td>
</tr>
<tr>
<td>Tuple</td>
<td>Row</td>
</tr>
<tr>
<td>Attribute</td>
<td>Column, or Field</td>
</tr>
</tbody>
</table>

3.1.2.2 Schema elements

A schema is made of

- Tables
- Type
- Domain
- View (result set of a stored query on the data)
- Assertion (constraints, transition constraints)
- Triggers (action to take after certain operations are performed)

Type and domains are two different things in some implementations, cf. for instance PostgreSQL, where a domain is defined to be essentially a datatype with constraint.10

3.1.2.3 Syntax

- A programming language: strict, cryptic error messages, tricky, evolves
- SQL is “kind of” case-insensitive11, doesn’t care about spaces and new lines
- Comments are with -- or /* ...*/
- Every statement ends with a ;
- Syntax is in Pb 3.1.
- We will stick to what’s in MariaDB and MySQL here (no domain, limited data type definition)

---

9For a clarification on the distinction between catalog and schemas, you can refer to e.g. https://stackoverflow.com/q/7022755.
11The SQL keywords are case-insensitive, but the table and schema names are sometimes case-sensitive, it depends on the actual implementation. For instance, MySQL is completely case-insensitive (reserved words, tables, attributes), MariaDB isn’t (case for table names matter).
3.2 First Commands

/* code/sql/HW_FACULTY.sql */

CREATE SCHEMA HW_FACULTY;

/* Or */
CREATE DATABASE HW_FACULTY;

CREATE TABLE HW_FACULTY.PROF(
  Fname VARCHAR(15), -- No String!
  Room INT, -- shorthand for INTEGER, are also available:
    SMALLINT, FLOAT, REAL, DEC
    -- The "REAL" datatype is like the "DOUBLE"
    -- datatype of C# (they are actually synonyms in
    -- SQL):
    -- more precise than the "FLOAT" datatype, but not
    -- as exact as the "NUMERIC" datatype.
    -- cf.
  Title CHAR(3), -- fixed-length string, padded with blanks if
    needed
  Tenured BIT(1),
  Nice BOOLEAN, -- True / False (= 0) / Unknown
  Hiring DATE, -- The DATE is always supposed to be entered in a
    YEAR/MONTH/DAY variation.
    -- To tune the way it will be displayed, you can
    -- use the "DATE_FORMAT" function
    -- (cf.
    -- but you can enter those values only using the
    "standard" literals
    -- (cf.
  Last_seen TIME,
  FavoriteFruit ENUM('apple','orange','pear'),
  PRIMARY KEY(Fname, Hiring)
);

/* Or, instead of using the fully qualified name HW_FACULTY.PROF, we could have: */
USE HW_FACULTY;
CREATE TABLE PROF(...)

/* */

USE HW_FACULTY;

INSERT INTO PROF VALUES ( 
3.3 Overview of Constraints

The following commands are particularly useful:

- `SHOW SCHEMAS;` -- List the schemas.
- `SHOW TABLES;` -- List the tables in a schema.
- `SHOW CREATE TABLE <TableName>;` -- Gives the command to "re-construct" TableName.
- `DESCRIBE <TableName>;` -- Show the structure of TableName.
- `SELECT * FROM <TableName>;` -- List all the rows in TableName.
- `DROP TABLE <TableName>;` -- "Drop" (erase) TableName.
- `DROP SCHEMA <SchemaName>;` -- "Drop" (erase) SchemaName.
- `SHOW WARNINGS;` -- After a warning was issued, show the content of the warning.

3.3 Overview of Constraints

1. Primary Key
2. Foreign Key
3. NOT NULL
4. UNIQUE
5. DEFAULT
6. CHECK

We know 1. and 2. from the Relational Model, here comes new constraints that cannot be describe in our relations.

```
CREATE TABLE HURRICANE(
    Name VARCHAR(25) PRIMARY KEY,
    WindSpeed INT DEFAULT 76,
    Above VARCHAR(25)
);
-- WindSpeed INT CHECK (WindSpeed > 74 AND WindSpeed < 500),
-- would be parsed but wouldn't have any effect!

CREATE TABLE STATE(
    Name VARCHAR(25) UNIQUE, -- NULL can be inserted
    Postal_abbr CHAR(2) NOT NULL
);
```

MySQL can output a description of those tables for us:
### 3.3 Overview of Constraints

MariaDB [HW_CONSTRAINTS_PART1]> DESCRIBE HURRICANE;
+----------+------------+------+-----+---------+-------+
| Field    | Type       | Null | Key | Default | Extra |
+----------+------------+------+-----+---------+-------+
| Name     | varchar(25) | NO   | PRI | NULL    |       |
| WindSpeed| int(11)    | YES  |     | 76      |       |
| Above    | varchar(25) | YES  | NULL|         |       |
+----------+------------+------+-----+---------+-------+
3 rows in set (0.01 sec)

MariaDB [HW_CONSTRAINTS_PART1]> DESCRIBE STATE;
+----------+------------+------+-----+---------+-------+
| Field    | Type       | Null | Key | Default | Extra |
+----------+------------+------+-----+---------+-------+
| Name     | varchar(25) | NO   | PRI | NULL    |       |
| Postal_abbr| char(2)   | NO   | UNI | NULL    |       |
+----------+------------+------+-----+---------+-------+
2 rows in set (0.00 sec)

--
-- Primary Keys
--

-- Adding a primary key:
ALTER TABLE STATE ADD PRIMARY KEY (Name);

-- Removing the primary key:
ALTER TABLE STATE DROP PRIMARY KEY;

--
-- UNIQUE Constraint
--

-- Adding a UNIQUE constraint
ALTER TABLE STATE ADD UNIQUE (Postal_abbr);

-- Removing a UNIQUE constraint
ALTER TABLE STATE DROP INDEX Postal_abbr;

--
-- NOT NULL Constraint
--

-- Adding the NOT NULL constraint
ALTER TABLE STATE MODIFY Postal_abbr CHAR(2) NOT NULL;

-- Removing the NOT NULL constraint
ALTER TABLE STATE MODIFY Postal_abbr CHAR(2);

--
3.3 Overview of Constraints

```
-- Default value
--

-- Changing the default value
ALTER TABLE HURRICANE ALTER COLUMN WindSpeed SET DEFAULT 74;

-- Removing the default value
ALTER TABLE HURRICANE ALTER COLUMN WindSpeed DROP DEFAULT;

-- Foreign key
--

-- Adding a foreign key constraint
ALTER TABLE HURRICANE ADD FOREIGN KEY (Above) REFERENCES STATE(Name);

-- Removing a foreign key constraint is out of the scope of this lecture.
-- If you are curious, you can have a look at
→ https://www.w3schools.com/sql/sql_foreignkey.asp
-- Dropping a foreign key constraint requires your constraint to have a
→ name, something we did not introduce.

• NOT NULL is to some extend part of the datatype.
• Note the difference between adding and removing the UNIQUE constraint: in one case, the
  parenthesis are mandatory, in the other, they would cause an error! In this example, the
  parenthesis around (Name) and (Postal_abbr) are mandatory.
• The datatype of the foreign key has to be the same as what we are referring.

A bit of testing:

INSERT INTO STATE VALUES('Georgia', 'GA');
INSERT INTO STATE VALUES('Texas', 'TX');
INSERT INTO STATE VALUES('FLORIDA', 'FL');
UPDATE STATE SET Name = 'Florida'
  WHERE Postal_abbr = 'FL';

-- There's an error with the following request. Why?
INSERT INTO HURRICANE VALUES('Irma', 150, 'FL');
/*
ERROR 1452 (23000): Cannot add or update a child row: a foreign key
  constraint fails ("HW_CONSTRAINTS_PART1"."HURRICANE", CONSTRAINT
  "HURRICANE_ibfk_1" FOREIGN KEY ("Above") REFERENCES "STATE" ("Name"))
*/

INSERT INTO HURRICANE VALUES('Harvey', DEFAULT, 'Texas');
INSERT INTO HURRICANE VALUES('Irma', 150, 'Florida');
DELETE FROM HURRICANE
  WHERE Name = 'Irma';
INSERT INTO HURRICANE VALUES('Irma', 150, 'Georgia');

UPDATE HURRICANE SET Above = 'Georgia'
```
WHERE Name = 'Irma';

/*
MariaDB [HW_CONSTRAINTS_PART1]> SELECT * FROM HURRICANE;
+--------+-----------+---------+
| Name   | WindSpeed | Above   |
|--------+-----------+---------|
| Harvey | 74        | Texas   |
| Irma   | 150       | Georgia |
+--------+-----------+---------+
*/

-- There's an error with the following request. Why?
UPDATE HURRICANE SET Above = 'North Carolina'
  WHERE Name = 'Irma';

-- Let's patch it, by adding North Carolina to our STATE table.
INSERT INTO STATE VALUES('North Carolina', 'NC');
UPDATE HURRICANE SET Above = 'North Carolina'
  WHERE Name = 'Irma';

3.4 Foreign Keys

3.4.1 A first example

CREATE TABLE STORM(
  Name VARCHAR(25) PRIMARY KEY,
  Kind ENUM('Tropical Storm', 'Hurricane'),
  WindSpeed INT,
  Creation DATE
);

-- I can change my enumerated datatype:
ALTER TABLE STORM MODIFY Kind ENUM('Tropical Storm', 'Hurricane',
  'Typhoon');

CREATE TABLE STATE(
  Name VARCHAR(25) UNIQUE,
  Postal_abbr CHAR(2) PRIMARY KEY,
  Affected_by VARCHAR(25),
  FOREIGN KEY (Affected_by) REFERENCES STORM(Name)
    ON DELETE SET NULL
    ON UPDATE CASCADE
);

INSERT INTO STORM VALUES('Harvey', 'Hurricane', 130, '2017-08-17');
-- In the following, the entry gets created, but date is 0000-00-00!
INSERT INTO STORM VALUES('Dummy', 'Hurricane', 120, '2017-08-08');
3.4 Foreign Keys

-- In the following, there's an error, and nothing gets inserted.

```
INSERT INTO STORM VALUES ('Dummy2', 'Hurricane', 120, DATE '2017-17-08');
```

-- The next one sets NULL for DATE.

```
INSERT INTO STORM VALUES ('Irma', 'Tropical Storm', 102, DEFAULT);
```

MySQL will always notify you if there is an error in a date attribute

```
INSERT INTO STATE VALUES ('Georgia', 'GA', NULL);
INSERT INTO STATE VALUES ('Texas', 'TX', NULL);
INSERT INTO STATE VALUES ('Florida', 'FL', NULL);
```

-- This instruction is not using the primary key, is that a problem?

```
UPDATE STATE SET Affected_by = 'Harvey'
    WHERE Name = 'Georgia';
```

```
UPDATE STORM SET Name = 'Harley' WHERE Name = 'Harvey';
DELETE FROM STORM
    WHERE Name = 'Harley';
```

### 3.4.2 Foreign keys restrictions

```
CREATE TABLE F_Key(
    Attribute VARCHAR(25) PRIMARY KEY
);
```

```
CREATE TABLE Table_default(
    Attribute1 VARCHAR(25) PRIMARY KEY,
    Attribute2 VARCHAR(25),
    FOREIGN KEY (Attribute2) REFERENCES F_Key(Attribute)
);
```

-- By default, this foreign key will restrict.

```
CREATE TABLE Table_restrict(
    Attribute1 VARCHAR(25) PRIMARY KEY,
    Attribute2 VARCHAR(25),
    FOREIGN KEY (Attribute2) REFERENCES F_Key(Attribute)
        ON DELETE RESTRICT
        ON UPDATE RESTRICT
);
```

```
CREATE TABLE Table_cascade(
    Attribute1 VARCHAR(25) PRIMARY KEY,
    Attribute2 VARCHAR(25),
    FOREIGN KEY (Attribute2) REFERENCES F_Key(Attribute)
        ON DELETE CASCADE
        ON UPDATE CASCADE
);
```
3.4 Foreign Keys

CREATE TABLE Table_set_null(
    Attribute1 VARCHAR(25) PRIMARY KEY,
    Attribute2 VARCHAR(25),
    FOREIGN KEY (Attribute2) REFERENCES F_Key(Attribute)
    ON DELETE SET NULL
    ON UPDATE SET NULL
);

/*
 * You might encounter a
 * ON UPDATE SET DEFAULT
 * but this reference option (cf.
 * worked only with a particular engine ( cf.
 * and won't be treated here.
*/

INSERT INTO F_Key VALUES('First Test');
INSERT INTO Table_default VALUES('Default', 'First Test');
INSERT INTO Table_restrict VALUES('Restrict', 'First Test');
INSERT INTO Table_cascade VALUES('Cascade', 'First Test');
INSERT INTO Table_set_null VALUES('Set null', 'First Test');

SELECT * FROM Table_default;
SELECT * FROM Table_restrict;
SELECT * FROM Table_cascade;
SELECT * FROM Table_set_null;

-- The following will fail because of the Table_default table:
UPDATE F_Key SET Attribute = 'After Update'
    WHERE Attribute = 'First Test';
DELETE FROM F_Key
    WHERE Attribute = 'First Test';

-- Let us drop this table, and try again.
DROP TABLE Table_default;

-- The following fails too, this time because of the Table_restrict table:
UPDATE F_Key SET Attribute = 'After Update'
    WHERE Attribute = 'First Test';
DELETE FROM F_Key
    WHERE Attribute = 'First Test';

-- Let us drop this table, and try again.
DROP TABLE Table_restrict;

-- Let's try again:
3.4 Foreign Keys

```sql
UPDATE F_Key SET Attribute = 'After Update' WHERE Attribute = 'First Test';

-- And let's print the situation after this update:
SELECT * FROM Table_cascade;
SELECT * FROM Table_set_null;

/

MariaDB [HW_CONSTRAINTS_PART3]> SELECT * FROM Table_cascade;
+------------+--------------+
<table>
<thead>
<tr>
<th>Attribute1</th>
<th>Attribute2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cascade</td>
<td>After Update</td>
</tr>
</tbody>
</table>
+------------+-------------+
1 row in set (0.00 sec)

MariaDB [HW_CONSTRAINTS_PART3]> SELECT * FROM Table_set_null;
+------------+------------+
<table>
<thead>
<tr>
<th>Attribute1</th>
<th>Attribute2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set null</td>
<td>NULL</td>
</tr>
</tbody>
</table>
+------------+------------+
1 row in set (0.00 sec)

*/

-- Let's make a second test.
INSERT INTO F_Key VALUES('Second Test');
INSERT INTO Table_cascade VALUES('Default', 'Second Test');
INSERT INTO Table_set_null VALUES('Restrict', 'Second Test');

DELETE FROM F_Key
WHERE Attribute = 'Second Test';

/

MariaDB [HW_CONSTRAINTS_PART3]> SELECT * FROM Table_cascade;
+------------+--------------+
<table>
<thead>
<tr>
<th>Attribute1</th>
<th>Attribute2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cascade</td>
<td>After Update</td>
</tr>
</tbody>
</table>
+------------+-------------+
1 row in set (0.00 sec)

MariaDB [HW_CONSTRAINTS_PART3]> SELECT * FROM Table_set_null;
+------------+------------+
<table>
<thead>
<tr>
<th>Attribute1</th>
<th>Attribute2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restrict</td>
<td>NULL</td>
</tr>
<tr>
<td>Set null</td>
<td>NULL</td>
</tr>
</tbody>
</table>
+------------+------------+
2 rows in set (0.00 sec)
```
3.4.3 Constructing and populating a new example

3.4.3.1 Construction

- Remember, we start by creating a schema and tables inside of it.
- What if foreign keys are mutually dependent? What if we have something like:

```sql
CREATE TABLE PROF(
    Login VARCHAR(25) PRIMARY KEY,
    Name VARCHAR(25),
    Department CHAR(5)
);

CREATE TABLE DEPARTMENT(
    Code CHAR(5) PRIMARY KEY,
    Name VARCHAR(25),
    Head VARCHAR(25),
    FOREIGN KEY (Head) REFERENCES PROF(Login)
        ON UPDATE CASCADE
);

ALTER TABLE PROF ADD FOREIGN KEY (Department)
    REFERENCES DEPARTMENT(Code);
```

Note the structure of the `ALTER TABLE` command:

- ... KEY Department REFERENCES Code;⇒ error
- ... KEY (Department) REFERENCES (Code);⇒ error
- ... KEY PROF(Department) REFERENCES DEPARTMENT(Code);⇒ ok

```sql
CREATE TABLE STUDENT(
    Login VARCHAR(25) PRIMARY KEY,
    Name VARCHAR(25),
    Registered DATE,
    Major CHAR(5),
    FOREIGN KEY (Major) REFERENCES DEPARTMENT(Code)
);

CREATE TABLE GRADE(
    Login VARCHAR(25),
```
3.5 A First Look at Conditions

Order of clauses does not matter, not even for optimization purpose.

```
UPDATE <table>
SET <attribute1> = <value1>, <attribute2> = <value2>, ...
WHERE <condition>;
```
3.6 Three-Valued Logic

Cf. (Elmasri and Navathe 2010, 5.1.1), (Elmasri and Navathe 2015, 7.1.1)
### 3.6.1 Meaning of NULL

NULL is

1. **Unknown value ("Nobody knows")**
   - What is the date of birth of Jack the Ripper?\(^\text{12}\)?
   - Does P equal NP?\(^\text{13}\)

2. **Unavailable / Withheld ("I don’t have that information with me at the moment")**
   - What is the number of English spies in France?
   - What is the VIN of your car?
   - What is the identity of the Tiananmen Square person?

3. **Not Applicable ("Your question doesn’t make sense")**
   - What is the US SSN of a French person?
   - What is the email address of an author of the XIXth century?

### 3.6.2 Comparison with unknown values \{truth_tables\}

If NULL is involved in a comparison, the result evaluates to “Unknown”.

<table>
<thead>
<tr>
<th>AND</th>
<th>T</th>
<th>F</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>U</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>U</td>
<td>U</td>
<td>F</td>
<td>U</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OR</th>
<th>T</th>
<th>F</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>U</td>
</tr>
<tr>
<td>U</td>
<td>T</td>
<td>U</td>
<td>U</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NOT</th>
<th>T</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>U</td>
<td></td>
</tr>
</tbody>
</table>

You can test if a value is NULL with `IS NULL`.

1. \texttt{INSERT INTO DEPARTMENT Values ('Hist', 'History', NULL);}  
2. \texttt{SELECT * FROM DEPARTMENT WHERE Head IS NULL;}  
3. \texttt{SELECT * FROM DEPARTMENT WHERE Head IS NOT NULL;}  
4. \texttt{SELECT COUNT(*) FROM GRADE WHERE Grade IS NULL;}  

There are no \texttt{if...then...else} statements in SQL, but you can do something similar with \texttt{CASE} (cf. 

\(^\text{12}\)https://en.wikipedia.org/wiki/Jack_the_Ripper  
\(^\text{13}\)https://en.wikipedia.org/wiki/List_of_unsolved_problems_in_computer_science
https://dev.mysql.com/doc/refman/8.0/en/case.html). However, note that SQL is probably not the right place to try to control the flow of execution.

Even with three times (!) the verbose option, MySQL does not give more insight as to whenever it drops comparing values once a NULL was encountered (cf. https://dev.mysql.com/doc/refman/8.0/en/mysql-command-options.html#option_mysql_verbose, you can log-in using `mysql -u testuser -p --password=password -v -v -v` to activate the most verbose mode).

Normally, `EXPLAIN` (https://dev.mysql.com/doc/refman/8.0/en/explain.html) should be useful in determining whenever MySQL uses some form of short-cut evaluation when comparing with `NULL`, but I could not make it work.

### 3.7 Various Tools

For `DISTINCT, ALL` and `UNION`, cf. (Elmasri and Navathe 2010, 4.3.4) or (Elmasri and Navathe 2015, 6.3.4). For `ORDER BY`, cf. (Elmasri and Navathe 2010, 4.3.6) or (Elmasri and Navathe 2015, 6.3.6). For aggregate functions, cf. (Elmasri and Navathe 2010, 5.1.7) or (Elmasri and Navathe 2015, 7.1.7).

### 3.8 AUTO_INCREMENT

Something that is not exactly a constraint, but that can be used to “qualify” domains, is the `AUTO_INCREMENT` feature of MySQL. Cf. https://dev.mysql.com/doc/refman/8.0/en/example-auto-increment.html, you can have MySQL increment a particular attribute (most probably intended to be your primary key) for you.

#### 3.8.1 Transactions

We can save the current state, and start a series of transactions, with the command

```
START TRANSACTION;
```

All the commands that follows are “virtually” executed: you can undo them all using

```
ROLLBACK;
```

This puts you back in the state you were in before starting the transaction. If you want all the commands you typed in-between to be actually enforced, you can use the command

```
COMMIT;
```

Nested transactions are technically possible, but they are counter-intuitive and should be avoided, cf. https://www.sqlskills.com/blogs/paul/a-sql-server-dba-myth-a-day-2630-nested-transactions-are-real/.

#### 3.8.2 DISTINCT / ALL

The result of a `SELECT` query, for instance, is a table, and SQL treats tables as multi-set, hence there can be repetitions in the result of a query, but we can remove them:

```
SELECT DISTINCT Major FROM STUDENT;
```

The default behaviour is equivalent to specifying `ALL`, and it display the duplicates.
3.8.3 UNION

( SELECT Login FROM STUDENT ) UNION ( SELECT Login FROM PROF );

There is also INTERSECT and EXCEPT in the specification, but MySQL does not implement them (cf. https://en.wikipedia.org/wiki/Comparison_of_relational_database_management_systems#Database_capabilities).

3.8.4 ORDER BY

You can have ORDER BY specifications:

SELECT Login FROM GRADE
  WHERE Grade > 3.0
  ORDER BY Grade;

SELECT Login FROM GRADE
  WHERE Grade > 3.0
  ORDER BY Grade DESC;

SELECT Login, Major FROM STUDENT
  ORDER BY Major, Name;

ORDER BY order by ascending order by default.

3.8.5 Aggregate functions

You can use MAX, SUM, MIN, AVG, COUNT:

SELECT MAX(Registered) FROM STUDENT;

SELECT COUNT(Name) FROM STUDENT;

SELECT COUNT(DISTINCT Name) FROM STUDENT;

3.8.6 Aliases for columns

SELECT Login FROM PROF;

+------------+
| Login      |
+------------+
| aturing    |
| caubert    |
| bgates     |
| perdos     |
+------------+

SELECT Login AS Username FROM PROF;

+------------+
| Username   |
+------------+
| aturing    |

+------------+
The purpose of aliases will be clearer as we study select-project-join queries.

### 3.9 More Select Queries

For select-project-join, cf. (Elmasri and Navathe 2010, 4.3.1) or (Elmasri and Navathe 2015, 6.3.1). For aliases, cf. (Elmasri and Navathe 2010, 4.3.2) or (Elmasri and Navathe 2015, 6.3.2). For nested queries, cf. (Elmasri and Navathe 2010, 5.1.2) or (Elmasri and Navathe 2015, 7.1.2).

#### 3.9.1 Select-project-join

```sql
SELECT Login FROM PROF, DEPARTMENT
WHERE DEPARTMENT.Name = 'Mathematics'
AND Department = Code;
```

- `Department.Name = 'Mathematics'` is the selection condition
- `Department = Code` is the join condition, because it combines two tuples.
- Why do we use the fully qualified name attribute for `Name`?
- We have to list all the tables we want to consult, even if we use fully qualified names.

```sql
SELECT Name FROM STUDENT, GRADE
WHERE Grade > 3.0
AND STUDENT.Login = GRADE.Login;
```

- `Grade > 3.0` is the selection condition
- `STUDENT.Login = GRADE/Login` is the join condition

We can have two join conditions:

```sql
SELECT PROF.Name FROM PROF, DEPARTMENT, STUDENT
WHERE STUDENT.Name = 'Ava Alyx'
AND STUDENT_MAJOR = DEPARTMENT_CODE
AND DEPARTMENT.Head = PROF.Login;
```

#### 3.9.2 Aliasing tuples

We can use aliases to shorten the previous query:

```sql
SELECT PROF.Name FROM PROF, DEPARTMENT, STUDENT AS B
WHERE B.Name = 'Ava Alyx'
AND B.MAJOR = DEPARTMENT_CODE
AND DEPARTMENT.Head = PROF.Login;
```

We can use multiple aliases:

```sql
SELECT A.Name FROM PROF AS A, DEPARTMENT, STUDENT AS B
WHERE B.Name = 'Ava Alyx'
AND B.MAJOR = DEPARTMENT_CODE
AND DEPARTMENT.Head = A.Login;
```
For those two, aliases were convenient, but not required to write the query. In some cases, we cannot do without aliases, for instance if we want to compare two rows in the same table:

```sql
SELECT Others.Login FROM GRADE AS Mine, GRADE AS Others
WHERE Mine/Login = 'aalyx'
AND Mine/Grade < Others/Grade;
```

AND NOT Me = Fellow would not work. Note that AS 'Fellow of Ava' is another kind of aliasing, mentioned in a previous section.

### 3.9.3 Nested queries

```sql
SELECT Login FROM GRADE
WHERE Grade > (SELECT AVG(Grade) FROM GRADE);
```

A nested query is made of an outer query (SELECT Login...) and an inner query (SELECT AVG(Grade)...).

(Average of all non-NULL values.)

```sql
SELECT Login FROM GRADE
WHERE Grade >= ALL (SELECT Grade FROM GRADE WHERE Grade IS NOT NULL);
```

Why cannot we use =?

Actually, nested query that uses = can often be rewritten without being nested:

```sql
SELECT Login
FROM PROF
WHERE DEPARTMENT = (SELECT Major
FROM STUDENT
WHERE Login LIKE 'a');
```

becomes

```sql
SELECT PROF/Login
FROM PROF, STUDENT
WHERE DEPARTMENT = Major AND STUDENT/Login = "cjoella";
```

Conversly, you can sometimes write select-project-join as nested queries For instance,
### 3.10 Setting Up Your Work Environment

This part is a short tutorial to install and configure a working relational DBMS.

#### 3.10.1 Installation

You will install the MySQL DataBase Management System, or its community-developed fork, MariaDB. Below are the instruction to install MySQL Community Edition on Windows 10, and MariaDB on Linux-based distribution, but both are developed for every major operating system (Mac OS, Windows, Debian, Ubuntu, etc.): feel free to pick one or the other, it won’t really make a difference in this course (up to some minor aspects). MySQL is more common, MariaDB is growing, both are released under GNU General Public License, well-documented and free of charge for their “community” versions.

It is perfectly acceptable to install MySQL or MariaDB on a virtual machine for this class. Remember that using your on the hub account, you should have access to VMware and Windows licences, and that using Virtual Box should also work fine. Once this step is completed, or if MySQL or MariaDB is already installed on your computer, go to the next step.

I tried to give precise and up-to-date instructions below, follow them carefully, read the messages displayed on your screen, make sure a step was correctly executed before moving to the next one, and everything should be all right. Also, remember:

1. Don’t wait, set your system early.
2. I will be happy to assist you, but there’s nothing I can do with an email like “It doesn’t work”: if you look for help, be detailed and clear about what you think went wrong.

The following links could be useful:


```sql
SELECT Name FROM STUDENT, GRADE
   WHERE Grade > 3.0
   AND STUDENT.Login = GRADE.Login;
```

becomes

```sql
SELECT Name FROM STUDENT
   WHERE Login IN (SELECT Login FROM GRADE WHERE Grade > 3.0);
```
3.10.1.1 Installing MySQL on Windows 10


where XXX is a number version (e.g., 8.0.13.0.), and YYY is the size of the file (e.g., 16.3M). On the next page, click on the (somewhat hidden) “No thanks, just start my download.” button.

2. Save the “mysql-installer-web-community-XXX.msi” file, and open it. If there is an updated version of the installer available, agree to download it. Accept the license term.

3. We will now install the various components needed for this class, leaving all the choices by defaults. This means that you need to do the following:

   a) Leave the first option on “Developer Default” and click on “Next”, or click on “Custom”, and select the following:

   b) Click on “Next” even if you don’t meet all the requirements

   c) Click on “Execute”. The system will download and install several softwares (this may take some time).

   d) Click on “Next” twice, leave “Type and Networking” on “Standalone MySQL Server / Classic MySQL Replication” and click “Next”, and leave the next options as they are (unless you know what you do and want to change the port, for instance) and click on “Next”.

   e) You now need to choose a password for the MySQL root account. It can be anything, just make sure to memorize it. Click on “Next”.

   f) On the “Windows Service” page, leave everything as it is and click on “Next”.

   g) On the “Plugins and Extensions” page, leave everything as it is and click on “Next”.

   h) Finally, click “Execute” on the “Apply Configuration” page, and then on “Finish”.

   i) Click on “Cancel” on the “Product Configuration” page and confirm that you don’t
want to add products: we only need to have MySQL Server XXX configured.

4. We now want to make sure that MySQL is running: launch Windows’ “Control Panel”, then click on “Administrative Tools”, and on “Services”. Look for “MySQLXX”, its status should be “Running”. If it is not, right-click on it and click on “Start”.

5. Open a command prompt (search for cmd, or use PowerShell\(^{22}\)) and type

   \texttt{cd \"C:\Program Files\MySQL\MySQL Server 8.0\bin\"}

   If this command fails, it is probably because the version number changed: open the file explorer, go to \texttt{C:\Program Files\MySQL\}, look for the right version number, and update the command accordingly.

   Then, enter

   \texttt{mysql -u root -p}

   and enter the password you picked previously for the root account. You are now logged as root in your database management system, you should see a brief message, followed by a prompt

   \texttt{mysql >}

3.10.1.2 Installing MySQL on macOS

The instructions are almost the same as for Windows. Read \url{https://dev.mysql.com/doc/refman/5.6/en/osx-installation-pkg.html} and download the file from \url{https://dev.mysql.com/downloads/mysql/} once you selected “macOS” as your operating system. Install it, leaving everything by default but adding a password (refer to the instructions for windows). Then, open a command-line interface (the terminal), enter

\texttt{mysql -u root -p}

and enter the password you picked previously for the root account. You are now logged as root in your database management system, you should see a brief message, followed by a prompt

\texttt{mysql >}

3.10.1.3 Installing MariaDB on Linux

1. Install, through your standard package management system (\texttt{apt} or \texttt{aptitude} for debian-based systems, \texttt{pacman} for Arch Linux, etc.), the packages \texttt{mysql-client} and \texttt{mysql-server} as well as their dependencies\(^{23}\).

2. Open a terminal and type

   \texttt{/etc/init.d/mysql status}

   to see if MySQL is running: if you read

   \texttt{Active: active (running)}

   then you can move on to the next step, otherwise run (as root)

\(^{22}\)\url{https://docs.microsoft.com/en-us/powershell/scripting/getting-started/getting-started-with-windows-powershell?view=powershell-6}

\(^{23}\)Yes, the package is called \texttt{mysql-server}, but it actually install the package \texttt{mariadb-server-10.1}... So don’t be confused: \textit{we are, indeed, installing MariaDB}!
3.10 Setting Up Your Work Environment

```bash
3. As root, type in your terminal

mysql_secure_installation
```

You’ll be asked to provide the current password for the root MySQL user: this password has not be defined yet, so just hit “Enter”. You’ll be asked if you want to set a new password (that you can freely chose, just make sure to memorize it). Then, answer “n” to the question “Remove anonymous users?”, “Y” to “Disallow root login remotely?”, “n” to “Remove test database and access to it?” and finally “Y” to “Reload privilege tables now?”.

4. Still as root, type in your terminal

```bash
mysql -u root -p
```

and enter the password you picked previously for the root account. You are now logged as root in your database management system: you should see a brief message, followed by a prompt

```
MariaDB [(none)]>
```

### 3.10.2 Creating an user

This step will create a non-root user and grant it some rights. Copy-and-paste or type the following three commands, one by one (that is, enter the first one, hit “enter”, enter the second, hit “enter”, etc.):

We first create a new user called `testuser` on our local installation, and give it the password `password`:

```
CREATE USER 'testuser'@'localhost' IDENTIFIED BY 'password';
```

Then, we grant it all the privileges on the databases whose name starts with HW_

```
GRANT ALL PRIVILEGES ON `HW_\%` . * TO 
'testuser'@'localhost';
```

Be careful: backticks (`) are surrounding HW\_\% whereas single quotes (') are surrounding testuser and localhost.

And then we quit the DBMS, using

```
EXIT;
```

The message displayed after the two first commands should be

```
Query OK, 0 rows affected (0.00 sec)
```

and the message displayed after the last command should be

```
Bye
```

---

24By default, MySQL and MariaDB only create a root user with all privileges and no password, but we added a password at the previous step.
3.10.3 Loging-in as testuser

We now log in as the normal user called “testuser”. Linux users should type as a normal user, i.e., not as root, in their terminal the following, and Windows users should type in their command prompt the following:

```
mysql -u testuser -p
```

Enter password as your password. If you are prompted with a message ERROR 1045 (28000): Access denied for user 'testuser'@'localhost' (using password: YES) then you probably typed the wrong password. Otherwise, you should see a welcoming message from MySQL or MariaDB and a prompt.

To save yourself the hassle of typing the password, you can use

```
mysql -u testuser -ppassword
```

or

```
mysql -u testuser -p --password=password
```

If at some point you want to know if you are logged as root or testuser, simply enter `\s;`

3.10.4 Creating our first database

Now, let us create our first schema, our first table, populate it with data, and display various information.

We first create the schema (or database), HW_FirstTest:

```
CREATE DATABASE HW_FirstTest; -- Or CREATE SCHEMA HW_FirstTest;
```

Let us make sure that we created it:

```
SHOW DATABASES;
```

Let us use it:

```
USE HW_FirstTest;
```

And see what it contains now:

```
SHOW TABLES;
```

We now create a table called TableTest, with two integer attributes called Attribute1 and Attribute2:

```
CREATE TABLE TableTest (Attribute1 INT, Attribute2 INT);
```

And can make sure that the table was indeed created:

```
SHOW TABLES;
```

---

25Provided the working directory is still C:\Program Files\MySQL\MySQL Server 5.7\bin. Cf. https://dev.mysql.com/doc/mysql-windows-excerpt/5.7/en/mysql-installation-windows-path.html to add the MySQL bin directory to your Windows system PATH environment variable.
We can further ask our DBMS to display the structure of the table we just created:

```
DESCRIBE TableTest; -- Can be abbreviated as DESC TableTest;
```

And even ask to get back the code that would create the exact same structure (but without the data!):

```
SHOW CREATE TABLE TableTest;
```

Now, let us populate it with some data:

```
INSERT INTO TableTest
VALUES (1,2),
       (3,4),
       (5,6);
```

Note that the SQL syntax and your DBMS are completely fine with your statement spreading over multiple lines. Let us now display the data stored in the table:

```
SELECT * FROM TableTest;
```

After that last command, you should see

```
+------------+------------+
| Attribute1 | Attribute2 |
+------------+------------+
| 1          | 2          |
| 3          | 4          |
| 5          | 6          |
+------------+------------+
```

Finally, we can erase the content of the table, then erase ("drop") the table, and finally the schema:

```
DELETE FROM TableTest; -- Delete the rows
DROP TABLE TableTest; -- Delete the table
DROP DATABASE HW_FirstTest; -- Delete the schema
```

You’re all set! All you have to do is to quit, using the command

```
EXIT;
```

**Exercises**

**Exercise 3.1** What does it mean to say that SQL is at the same time a “data definition language” and a “data manipulation language”?

**Exercise 3.2** Name three kind of objects (for lack of a better word) a `CREATE` statement can create.

**Exercise 3.3** Write a SQL statement that adds a primary key constraint to an attribute named `Id` in an already existing table named `STAFF`.

**Exercise 3.4** Complete each row of the following table with either a datatype or two different examples:

<table>
<thead>
<tr>
<th>Data type</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4, -32</td>
</tr>
</tbody>
</table>
### Exercise 3.5
In the datatype CHAR(3), what does the 3 indicate?

### Exercise 3.6
Explain this query: `CREATE SCHEMA FACULTY;`

### Exercise 3.7
Explain this query:

```sql
ALTER TABLE TABLEA
DROP INDEX Attribute1;
```

### Exercise 3.8
If I want to enter January 21, 2016, as a value for an attribute with the DATE datatype, what value should I enter?

### Exercise 3.9
Write a statement that inserts the values "Thomas" and 4 into the table TRAINS.

### Exercise 3.10
If PkgName is the primary key in the table MYTABLE, what can you tell about the number of rows returned by the following statement?

```sql
SELECT * FROM MYTABLE WHERE PkgName = 'MySQL';
```

### Exercise 3.11
If you want that every time a referenced row is deleted, all the referring rows are deleted as well, what mechanism should you use?

### Exercise 3.12
If a database designer is using the ON UPDATE SET NULL for a foreign key, what mechanism is (s)he implementing (i.e., describe how the database will react a certain operation)?

### Exercise 3.13
If the following is part of the design of a table:

```sql
FOREIGN KEY (DptNumber) REFERENCES DEPARTMENT(Number)
ON DELETE SET DEFAULT
ON UPDATE CASCADE;
```

What happen to the rows whose foreign key DptNumber are set to 3 if the row in the DEPARTEMENT table with primary key Number set to 3 is...

1. ... deleted?
2. ...updated to 5?

### Exercise 3.14
If the following is part of the design of a WORKER table:

```sql
FOREIGN KEY WORKER(DptNumber) REFERENCES DEPARTMENT(DptNumber)
ON UPDATE CASCADE;
```

What happen to the rows whose foreign key DptNumber are set to 3 if the row in the DEPARTMENT table with primary key Number set to 3 is...

1. ... deleted?
2. ... updated to 5?

### Exercise 3.15
Given a relation TOURIST(Name, EntryDate, Address), write a SQL statement printing the name and address of all the tourists who entered the territory after the 15 September, 2012.
Exercise 3.16  Describe what the star do in the statement

```
SELECT ALL * FROM MYTABLE;
```  

Exercise 3.17  What is the fully qualified name of an attribute? Give an example.

Exercise 3.18  If DEPARTMENT is a database, what is DEPARTMENT.*?

Exercise 3.19  What is a multi-set? What does it mean to say that SQL treats tables as multisets?

Exercise 3.20  What is the difference between

```
SELECT ALL * FROM MYTABLE;
```

and

```
SELECT DISTINCT * FROM MYTABLE;
```

How are the results the same? How are they different?

Exercise 3.21  What is wrong with the statement

```
SELECT * WHERE Name = 'CS' FROM DEPARTMENT;
```

Exercise 3.22  Write a query that returns the number of row (i.e., of entries, of tuples) in a table named BOOK.

Exercise 3.23  When is it useful to use a select-project-join query?

Exercise 3.24  When is a tuple variable useful?

Exercise 3.25  Write a query that changes the name of the professor whose Login is 'caubert' to 'Hugo Pernot' in the table PROF.

Exercise 3.26  Can an UPDATE statement have a WHERE condition using an attribute that isn’t the primary key? If no, justify, if yes, tell what could happen.

Exercise 3.27  What are the possible meanings or interpretations for a NULL value?

Exercise 3.28  What are the values of the following expressions (i.e., do they evaluate to TRUE, FALSE, or UNKNOWN)?

- TRUE AND FALSE
- TRUE AND UNKNOWN
- NOT UNKNOWN
- FALSE OR UNKNOWN

Exercise 3.29  Write the truth table for AND for the three-valued logic of SQL.

Exercise 3.30  What comparison expression should you use to test if a value is different from NULL?

Exercise 3.31  Explain this query:

```
SELECT Login
FROM PROF
WHERE Department IN ( SELECT Major
                        FROM STUDENT
                        WHERE Login = 'jrakesh');
```

Can you rewrite it without nesting queries?
Exercise 3.32  What is wrong with this query?

```sql
SELECT Name FROM STUDENT
WHERE Login IN
  ( SELECT Code FROM Department WHERE head = 'aturing');
```

Exercise 3.33  Write a query that returns the sum of all the values stored in the Pages attribute of a BOOK table.

Exercise 3.34  Write a query that adds a Pages attribute of type INT into a (already existing) BOOK table.

Exercise 3.35  Write a query that removes the default value for a Pages attribute in a BOOK table.

Exercise 3.36  Under which conditions does SQL allow you to enter the same row in a table twice?

Exercises

Solution to Exercises

Solution 3.1  It can specify the conceptual and internal schema, and it can manipulate the data.

Solution 3.2  Database (schema), table, view, assertion, trigger, etc.

Solution 3.3  ```sql
ALTER TABLE STAFF ADD PRIMARY KEY(Id);
``` 

Solution 3.4  

<table>
<thead>
<tr>
<th>Data type</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int</td>
<td>4, -32</td>
</tr>
<tr>
<td>Char(4)</td>
<td>'trai', 'plol'</td>
</tr>
<tr>
<td>VarChar(10)</td>
<td>'Train', 'Michelle'</td>
</tr>
<tr>
<td>Bit(4)</td>
<td>B'1010', B'0101'</td>
</tr>
<tr>
<td>Boolean</td>
<td>TRUE, UNKNOWN</td>
</tr>
</tbody>
</table>

Solution 3.5  That we can store exactly three characters.

Solution 3.6  It creates a schema, i.e., a database, named Faculty.

Solution 3.7  It removes the UNIQUE constraint on the Attribute1 in the TABLEA table.

Solution 3.8  ```sql
DATE '2016-01-21', '2016-01-21', '2016/01/21', '20160121'.
``` 

Solution 3.9  ```sql
INSERT INTO TRAINS VALUES('Thomas', 4);
``` 

Solution 3.10  Yes.

Solution 3.11  We should use a referential triggered action clause, : ON DELETE CASCADE.

Solution 3.12  If the referenced row is updated, then the attribute of the referencing rows are set to NULL.

Solution 3.13  In the referencing rows,

1. the department number is set to the default value.
2. the department number is updated accordingly.
Solution 3.14  1. This operation is rejected: the row in the DEPARTMENT table with primary key Number set to 3 cannot be deleted if a row in the WORKER table references it.
   2. In the referencing rows, the department number is updated accordingly.

Solution 3.15  We could use the following:

```sql
SELECT Name, Address
FROM TOURIST
WHERE EntryDate > DATE'2012-09-15';
```

Solution 3.16  It selects all the attributes, it is a wildcard.

Solution 3.17  The name of the relation with the name of its schema and a period beforehand. An example would be EMPLOYEE. Name.

Solution 3.18  All the tables in that database.

Solution 3.19  A set where the same value can occur twice. The same row can occur twice in a table.

Solution 3.20  They both select all the rows in the MYTABLE table, but ALL will print the duplicate values, whereas DISTINCT will print them only once.

Solution 3.21  You cannot have the WHERE before FROM.

Solution 3.22  `SELECT COUNT(*) FROM BOOK;`

Solution 3.23  We use those query that projects on attributes using a selection and join conditions when we need to construct for information based on pieces of data spread in multiple tables.

Solution 3.24  It makes the distinction between two different rows of the same table, it is useful when we want to select a tuple in a relation that is in a particular relation with a tuple in the same relation. Quoting https://stackoverflow.com/a/7698796/:

They are useful for saving typing, but there are other reasons to use them:

- If you join a table to itself you must give it two different names otherwise referencing the table would be ambiguous.
- It can be useful to give names to derived tables, and in some database systems it is required... even if you never refer to the name.

Solution 3.25  We could use the following:

```sql
UPDATE PROF SET Name = 'Hugo Pernot'
WHERE Login = 'caubert';
```

Solution 3.26  Yes, we can have select condition that does not use primary key. In that case, it could be the case that we update more than one tuple with such a command (which is not necessarily a bad thing).

Solution 3.27  Unknown value, unavailable / withheld, N/A.

Solution 3.28  • TRUE AND FALSE → FALSE
   • TRUE AND UNKNOWN → UNKNOWN
   • NOT UNKNOWN → UNKNOWN
   • FALSE OR UNKNOWN → FALSE

Solution 3.29  • TRUE AND TRUE → TRUE
3.10 Setting Up Your Work Environment

- TRUE AND FALSE → FALSE
- TRUE AND UNKNOWN → UNKNOWN
- FALSE AND FALSE → FALSE
- UNKNOWN AND UNKNOWN → UNKNOWN
- FALSE AND UNKNOWN → FALSE
- The other cases can be deduced by symmetry.

For a more compact presentation, refer to (3.29truth_tables).

Solution 3.30 IS NOT

Solution 3.31 It lists the login of the professors teaching in the department where the student whose login is “jrakesh” is majoring. It can be rewritten as

```sql
SELECT PROF.Login
FROM PROF, STUDENT
WHERE Department = Major
AND STUDENT.Login = 'jrakesh';
```

Solution 3.32 It tries to find a Login in a Code.

Solution 3.33 SELECT SUM(Pages) FROM BOOK;

Solution 3.34 ALTER TABLE BOOK ADD COLUMN Pages INT;

Solution 3.35 ALTER TABLE BOOK ALTER COLUMN Pages DROP DEFAULT;

Solution 3.36 Essentially, if there are no primary key in the relation, and if no attribute has the UNIQUE constraint. Cf. also Solution 3.36 ().

Problems

Problem 3.1 (Discovering the documentation) The goal of this problem is to learn where to find the documentation for your DBMS, and to understand how to read the syntax of SQL commands.

You can consult your textbook, (Elmasri and Navathe 2010, Table 5.2, p. 140) or (Elmasri and Navathe 2015, Table 7.2, p. 235), for a very quick summary of the most common commands. Make sure you are familiar with the Backus–Naur form (BNF) notation commonly used:

- non-terminal symbols (i.e., variables, parameters) are enclosed in angled brackets, `<…>`
- optional parts are shown in square brackets, `[…]`
- repetitions are shown in braces `{…}`
- alternatives are shown in parenthesis and separated by vertical bars, `(…|…|…)`

The most complete lists of commands are probably at

- https://mariadb.com/kb/en/library/sql-statements/ and

Those are the commands implemented in the DBMS you are actually using. Since there are small variations from one implementation to the other, it is better to take one of this link as a reference in the future.

---

**Problem 3.2 (Creating and using a simple table in SQL)** This problem will guide you in manipulating a very simple table in SQL.

**Pb 3.2 – Question 1** Log in as testuser, create a database named HW_Address, use it, and create two tables:

```sql
CREATE TABLE NAME(
    FName VARCHAR(15),
    LName VARCHAR(15),
    Id INT,
    PRIMARY KEY(Id)
);
CREATE TABLE ADDRESS(
    StreetName VARCHAR(15),
    Number INT,
    Habitant INT,
    PRIMARY KEY(StreetName, Number)
);
```

**Pb 3.2 – Question 2** Observe the output produced by the command `DESC ADDRESS`.

**Pb 3.2 – Question 3** Add a foreign key to the ADDRESS table, using

```sql
ALTER TABLE ADDRESS
ADD FOREIGN KEY (Habitants)
REFERENCES NAME(Id);
```

And observe the new output produced by the command `DESC ADDRESS`.

Is it what you would have expected? How informative is it? Can you think of a command that would output more detailed information, including a reference to the existence of the foreign key?

**Pb 3.2 – Question 4** Draw the relational model corresponding to that database, including the primary, as well as foreign, keys.

**Pb 3.2 – Question 5** Add some data in the NAME table:

```sql
INSERT INTO NAME VALUES ('Barbara', 'Liskov', 003);
INSERT INTO NAME VALUES ('Tuong Lu', 'Kim', 004);
INSERT INTO NAME VALUES ('Samantha', NULL, 080);
```

What command can you use to display this information back? Do you notice anything regarding the values we entered for the Id attribute?

**Pb 3.2 – Question 6** Add some data into the ADDRESS table:

```sql
INSERT INTO ADDRESS VALUES
```
Problem 3.3 (Duplicate rows in SQL) Log in as testuser and create a database HW_REPETITION. Create in that database a table (I will assume you named it EXAMPLE in the following, but you are free to name it the way you want) with at least two attributes that have different data types. Don’t declare a primary key yet. Answer the following:

Pb 3.3 – Question 1 Add a tuple to your table using

\[
\text{INSERT INTO EXAMPLE VALUES}(X, Y);
\]

where X and Y are values with the right datatype. Try to add this tuple again. What do you observe? (You can use \text{SELECT * FROM} EXAMPLE; to observe what is stored in this table.)

Pb 3.3 – Question 2 Alter your table to add a primary key, using

\[
\text{ALTER TABLE EXAMPLE ADD PRIMARY KEY (Attribute)};
\]

where Attribute is the name of the attribute you want to be a primary key. What do you observe?

Pb 3.3 – Question 3 Empty your table using

\[
\text{DELETE FROM EXAMPLE};
\]

and alter your table to add a primary key, using the command we gave at the previous step. What do you observe?

Pb 3.3 – Question 4 Try to add the same tuple twice. What do you observe?

Problem 3.4 (Revisiting the PROF table) Create the PROF, STUDENT, DEPARTMENT and GRADE tables as in the “Constructing and populating a new example” section. Populate them with some data (copy it from the notes or invent it, it won’t matter).

Pb 3.4 – Question 1 Draw the complete relational model for this database (i.e., for the PROF, DEPARTMENT, STUDENT and GRADE relations).

Pb 3.4 – Question 2 Create and populate a LECTURE table as follows:
• It should have four attributes, Name, Instructor, Code, and Year, of type VARCHAR(25) for the two first, CHAR(5), and YEAR(4).
• The Year and Code attributes should be the primary key (yes, have two attributes be the primary key).
• The Instructor attribute should be a foreign key referencing the Login attribute in PROF.
• Populate the LECTURE table with some made-up data.

Try to think about some of the weaknesses of this representation (for instance, can it accommodate two instructors for the same class?), and write down two possible scenarios where this schema would not be appropriate.

Pb 3.4 – Question 3 The GRADE table had some limitations, too, since every student could have only one grade. Add two columns to the GRADE table, using

```sql
ALTER TABLE GRADE
ADD COLUMN LectureCode CHAR(5),
ADD COLUMN LectureYear YEAR(4);
```

and add a foreign key:

```sql
ALTER TABLE GRADE
ADD FOREIGN KEY (LectureYear, LectureCode)
REFERENCES LECTURE(Year, Code);
```

Use DESCRIBE and SELECT to observe the schema of the GRADE table and its rows. Is that what you would have expected?

Pb 3.4 – Question 4 Update the tuples in GRADE with some made-up data: now, every row should contain, on top of a login and a grade, a lecture year and a lecture code.

Pb 3.4 – Question 5 Update the relational model you previously drawn to reflect the new situation of your tables.

Pb 3.4 – Question 6

Write SELECT statements answering the following questions (where PROF.Name, LECTURE.Name, YYYY, LECTURE.Code and STUDENT.Login should be relevant values considering your data):

1. “Could you give me the logins and grades of the students who took LECTURE.Name in ‘YYYY’?”
2. “Could you list the instructors who taught in year YYYY?” (and, please, avoid duplicates)
3. “Could your list the name and grade of all the student who took class LECTURE.Code (no matter the year)?”
4. “Could you tell me which years was the class LECTURE.Code taught?”
5. “Could you list the classes taught the same year as class LECTURE.Code?”
6. “Could you print the name of the students who registered after STUDENT>Login?”
7. “How many departments’ heads are teaching this year?”

Problem 3.5 (TRAIN table and more advanced SQL coding) Look at the SQL code below, and then answer the following questions.
Pb 3.5 – Question 1 Modify the CREATE statement that creates the TRAIN table (l. 1–5), so that Id would be declared as the primary key. You can write only the line(s) that need to change.

Pb 3.5 – Question 2 Write an ALTER statement that makes Id become the primary key of the CONDUCTOR table.

Pb 3.5 – Question 3 Modify the CREATE statement that creates the ASSIGNED TO table (l. 13–18), so that it has two foreign keys: ConductorId would be referencing the Id attribute in CONDUCTOR and TrainId would be referencing the Id attribute in TRAIN. You can write only the line(s) that need to change.

Pb 3.5 – Question 4 Write INSERT statements that insert one tuple of your invention in each relation (without NULL values). Your statements should respect all the constraints (including the ones we added at the previous questions) and result in actual insertions. (Remember that four digits is a valid value for an attribute with the YEAR(4) datatype.)

Pb 3.5 – Question 5 Write a statement that sets the value of the ExperienceLevel attribute to “Senior” in all the tuples where the Id attribute is “GP1029” in the CONDUCTOR relation.

Pb 3.5 – Question 6 For each of the following questions, write a SELECT statement that would answer it:

1. "What are the identification numbers of the trains?"
2. "What are the names of the conductors with a "Senior" experience level?"
3. "What are the construction years of the "Surfliner" and "Regina" models that we have?"
4. "What is the id of the conductor that was responsible of the train referenced "K-13" on 2015/12/14?"
5. "What are the models that were ever conducted by the conductor whose id is "GP1029"?"
Problem 3.6 *(Insert, Select and Project in the COFFEE database)* Suppose we have the relational model depicted below, with the indicated data in it:

**COFFEE**

<table>
<thead>
<tr>
<th>Ref</th>
<th>Origin</th>
<th>TypeOfRoast</th>
<th>PricePerPound</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>Brazil</td>
<td>Light</td>
<td>8.90</td>
</tr>
<tr>
<td>121</td>
<td>Bolivia</td>
<td>Dark</td>
<td>7.50</td>
</tr>
<tr>
<td>311</td>
<td>Brazil</td>
<td>Medium</td>
<td>9.00</td>
</tr>
<tr>
<td>221</td>
<td>Sumatra</td>
<td>Dark</td>
<td>10.25</td>
</tr>
</tbody>
</table>

**CUSTOMER**

<table>
<thead>
<tr>
<th>CardNo</th>
<th>Name</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>Bob Hill</td>
<td><a href="mailto:b.hill@isp.net">b.hill@isp.net</a></td>
</tr>
<tr>
<td>002</td>
<td>Ana Swamp</td>
<td><a href="mailto:swampa@nca.edu">swampa@nca.edu</a></td>
</tr>
<tr>
<td>003</td>
<td>Mary Sea</td>
<td><a href="mailto:brig@gsu.gov">brig@gsu.gov</a></td>
</tr>
<tr>
<td>004</td>
<td>Pat Mount</td>
<td><a href="mailto:pmount@fai.fr">pmount@fai.fr</a></td>
</tr>
</tbody>
</table>

**SUPPLY**

<table>
<thead>
<tr>
<th>Provider</th>
<th>Coffee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee Unl.</td>
<td>001</td>
</tr>
<tr>
<td>Coffee Unl.</td>
<td>121</td>
</tr>
<tr>
<td>Coffee Exp.</td>
<td>311</td>
</tr>
<tr>
<td>Johns &amp; Co.</td>
<td>221</td>
</tr>
</tbody>
</table>

**PROVIDER**

<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee Unl.</td>
<td><a href="mailto:bob@cofunl.com">bob@cofunl.com</a></td>
</tr>
<tr>
<td>Coffee Exp.</td>
<td><a href="mailto:pat@coffeex.dk">pat@coffeex.dk</a></td>
</tr>
<tr>
<td>Johns &amp; Co.</td>
<td>NULL</td>
</tr>
</tbody>
</table>

In the following, we will assume that this model was implemented in a DBMS (MySQL or MariaDB), the primary keys being **COFFEE.Ref**, **CUSTOMER.CardNo**, **SUPPLY.Provider** and **SUPPLY.Coffee**, and **PROVIDER.Name**, and the foreign keys being as follow:

- **FavCoffee** in the **CUSTOMER** relation refers to **Ref** in the **COFFEE** relation
- **Provider** in the **SUPPLY** refers to **Name** in the **PROVIDER** relation
- **Coffee** in the **SUPPLY** refers to **Ref** in the **COFFEE** relation

You will be asked to read and write SQL commands. You should assume that

1. Datatype doesn’t matter: we use only strings and appropriate numerical datatypes.
2. Every statement respects SQL’s syntax (there’s no “a semi-colon is missing” trap).
3. None of those commands are actually executed: the data is always in the state depicted above, you are asked to answer “what if” questions.

You can use COFFEE.1 to denote the first tuple (or row) in COFFEE, and similarly for other relations and tuples (so that, for instance SUPPLY.4 corresponds to "Johns & Co"., 221).

Pb 3.6 – Question 1 Draw the relational model of this table.

Pb 3.6 – Question 2 Determine if the following insertion statements would violate the the entity integrity constraint, ("primary key cannot be NULL and should be unique), the referential integrity constraint ("the foreign key must refer to something that exists"), if there would be some other kind of error (ignoring the plausability / revelance of inserting that tuple), or if it would result in successful insertion.

```sql
1. INSERT INTO CUSTOMER VALUES(005, 'Bob Hill', NULL, 001);
2. INSERT INTO COFFEE VALUES(002, "Peru", "Decaf", 3.00);
3. INSERT INTO PROVIDER VALUES(NULL, "contact@localcof.com");
4. INSERT INTO SUPPLY VALUES("Johns Co.", 121);
5. INSERT INTO SUPPLY VALUES("Coffee Unl.", 311, 221);
```

Pb 3.6 – Question 3 Assuming that the referential triggered action clause ON UPDATE CASCADE is used for every foreign keys in this database, list the tuples modified by the following statements:

```sql
1. UPDATE CUSTOMER SET FavCoffee = 001
   WHERE CardNo = 001;
2. UPDATE COFFEE SET TypeOfRoast = 'Decaf'
   WHERE Origin = 'Brazil';
3. UPDATE PROVIDER SET Name = 'Coffee Unlimited'
   WHERE Name = 'Coffee Unl.';
4. UPDATE COFFEE SET PricePerPound = 10.00
   WHERE PricePerPound > 10.00;
```

Pb 3.6 – Question 4 Assuming that the referential triggered action clause ON DELETE CASCADE is used for every foreign keys in this database, list the tuples modified by the following statements:

```sql
1. DELETE FROM CUSTOMER
   WHERE Name LIKE '%S%';
2. DELETE FROM COFFEE
   WHERE Ref = 001;
3. DELETE FROM SUPPLY
   WHERE Provider = 'Coffee Unl.'
   AND Coffee = 001;
4. DELETE FROM PROVIDER
   WHERE Name = 'Johns & Co.';
```
3.10 Setting Up Your Work Environment

Pb 3.6 – Question 5 Starting here, assume that there is more data in our table than what was given at the beginning of the problem. Write queries that answer the following questions:

1. “What are the origins of your dark coffees?”
2. “What is the reference of Bob's favorite coffee?” (nota: it doesn’t matter if you return the favorite coffee of all the Bobs in the database.)
3. “What are the names of the providers who didn’t give their email?”
4. “How many coffees does Johns & co. provide us with?”
5. “What are the names of the providers of my dark coffees?”

Problem 3.7 (Select queries for the COMPUTER table) Consider the following SQL code:

```sql
/* code/sql/COMPUTER.sql */
CREATE TABLE COMPUTER(
    Id VARCHAR(20) PRIMARY KEY,
    Model VARCHAR(20)
);

CREATE TABLE PRINTER(
    Id VARCHAR(20) PRIMARY KEY,
    Model VARCHAR(20)
);

CREATE TABLE CONNEXION(
    Computer VARCHAR(20),
    Printer VARCHAR(20),
    PRIMARY KEY(Computer, Printer),
    FOREIGN KEY (Computer) REFERENCES COMPUTER(Id),
    FOREIGN KEY (Printer) REFERENCES PRINTER(Id)
);

INSERT INTO COMPUTER VALUES
    ('A', 'DELL A'),
    ('B', 'HP X'),
    ('C', 'ZEPTO D'),
    ('D', 'MAC Y');

INSERT INTO PRINTER VALUES
    ('12', 'HP-140'),
    ('13', 'HP-139'),
    ('14', 'HP-140'),
    ('15', 'HP-139');

INSERT INTO CONNEXION VALUES
    ('A', '12'),
    ('A', '13'),
    ('B', '13'),
    ('C', '14');
```
Write a query that returns ... (in parenthesis, the values returned in this set-up, but you have to be general)

1. ... the number of computers connected to the printer whose Id is 13 (i.e., 2).
2. ... the number of different models of printers (i.e., 2).
3. ... the model(s) of the printer connected to the computer whose Id is 'A' (i.e., 'HP-140' and 'HP-139').
4. ... the id of the computer(s) not connected to any printer (i.e., 'D').

Problem 3.8 (Improving a role-playing game) A friend of yours want you to review and improve the code for a role-playing game.

The original idea was that each character should have a name, a class (e.g., Bard, Assassin, Druid), a certain amount of experience, a level, one or more weapons (providing bonuses) and can complete quests. A quest have a name, and rewards the characters that completed it with a certain amount of experience, and sometimes (but rarely) with a special item.

Your friend came up with the following code:

```sql
CREATE TABLE CHARACTER(
    Name VARCHAR(30) PRIMARY KEY,
    Class VARCHAR(30),
    XP INT,
    LVL INT,
    Weapon_Name VARCHAR(30),
    Weapon_Bonus INT,
    Quest_Completed VARCHAR(30)
);

CREATE TABLE QUEST(
    Id VARCHAR(20) PRIMARY KEY,
    Completed_By VARCHAR(30),
    XP_Gained INT,
    Special_Item VARCHAR(20),
    FOREIGN KEY (Completed_By) REFERENCES CHARACTER(Name)
);

ALTER TABLE CHARACTER
    ADD FOREIGN KEY (Quest_Completed) REFERENCES QUEST(Id);
```

But there are several problems.

- As of now, a character can have only one weapon. All the attempts to “hack” the CHARACTER table to add an arbitrary number of weapons ended up creating horrible messes.
- Every time a character completes a quest, a copy of the quest must be created. Your friend is not so sure why, but nothing else works. Also it seems that a character can complete only one quest, but your friend is not so sure about that.
- It would be nice to be able to store features that are tied to the class, and not to the character, like the bonus they provide and the associated element (e.g., all bards use fire, all assassins use wind, etc.). But you friend simply cannot figure out how to do that.
Can you provide a relational model (no need to write the SQL code, but remember to indicate the primary and foreign keys) that would solve all of your friend’s troubles?

**Problem 3.9 (A simple database for books)** Consider the following code:

```sql
/*

code/sql/SIMPLE_BOOK.sql

The following is in case you want to run your program on your installation:

-- DROP SCHEMA HW_SIMPLE_BOOK;
CREATE SCHEMA HW_SIMPLE_BOOK;
USE HW_SIMPLE_BOOK;

CREATE TABLE AUTHOR(
    FName VARCHAR(30),
    LName VARCHAR(30),
    Id INT PRIMARY KEY
);

CREATE TABLE PUBLISHER(
    Name VARCHAR(30),
    City VARCHAR(30),
    PRIMARY KEY (Name, City)
);

CREATE TABLE BOOK(
    Title VARCHAR(30),
    Pages INT,
    Published DATE,
    PublisherName VARCHAR(30),
    PublisherCity VARCHAR(30),
    FOREIGN KEY (PublisherName, PublisherCity)
        REFERENCES PUBLISHER(Name, City),
    Author INT,
    FOREIGN KEY (Author)
        REFERENCES AUTHOR(Id),
    PRIMARY KEY (Title, Published)
);

INSERT INTO AUTHOR VALUES
    ("Virginia", "Wolve", 01),
    ("Paul", "Bryant", 02),
    ("Samantha", "Carey", 03);

INSERT INTO PUBLISHER VALUES
```
The values inserted in the database will guide some examples, but you should assume there is more data than what we inserted. In this long problem, you will be asked to write commands to select, update, delete, insert data, and to enhance the model.

1. Write a command that selects...
   a) the title of all the books.
   b) the (distinct) name of the publishers.
   c) the title and publication date of the books published since January 31, 2012.
   d) the first and last names of the authors published by Gallimard (no matter the city).
   e) the first and last names of the authors who were not published by an editor in New-York.
   f) the id of the authors who published a book whose name starts with "Where".
   g) the total number of pages in the database.
   h) the number of pages in the longest book written by the author whose last name is "Wolve".
   i) the title of the books published in the XIXth century.

2. Write a command that updates the title of all the books written by the author whose id is 1 to "BANNED". Is there any reason for this command to be rejected by the system? If yes, explain which one.

3. Write one or multiple commands that would delete the author whose id is 3, and all the book written by that author. Make sure you don’t violate any foreign key constraint.

4. Write a command that would create a table used to record the awards granted to authors for particular books. You should assume that each award has its own name, is awarded every year, and that it is awarded to an author for a particular book. Pick appropriate attributes, datatypes, primary as well as foreign keys, and avoid above all redundancy.

5. Draw the relational model of the database you created (i.e., including all the relations given in the code and the one you added).

6. Discuss two limitations of the model and how to improve it.

---

26 You can use the DATE datatype to store a year.
Solution to Selected Problems

Solution to Problem 3.2 (Creating and using a simple table in SQL) This problem is supposed to be a straightforward application of what we studied in class. Look back at Setting Up Your Work Environment if you feel like you are stuck before rushing to this solution.

Pb 3.2 – Solution to Q. 1 We simply log-in as indicated in the “Log-in as testuser” Section. Then, we enter

```sql
1. CREATE DATABASE HW_Address;
2. USE HW_Address;
```

We then create the tables as in the problem.

Pb 3.2 – Solution to Q. 2 Omitting the Extra column, we have:

```
MariaDB [HW_Address]> DESC ADDRESS;
+---------------+--------------+------+-----+--------+
| Field         | Type         | Null | Key | Default|
+---------------+--------------+------+-----+--------+
| StreetName    | varchar(15)  | NO   | PRI | NULL   |
| Number        | int(11)      | NO   | PRI | NULL   |
| Habitants     | int(11)      | YES  |     | NULL   |
+---------------+--------------+------+-----+--------+
```

Pb 3.2 – Solution to Q. 3 We add the foreign key, then (still omitting the Extra column):

```
MariaDB [HW_Address]> DESC ADDRESS;
+---------------+--------------+------+-----+--------+
| Field         | Type         | Null | Key | Default|
+---------------+--------------+------+-----+--------+
| StreetName    | varchar(15)  | NO   | PRI | NULL   |
| Number        | int(11)      | NO   | PRI | NULL   |
| Habitants     | int(11)      | YES  | MUL | NULL   |
+---------------+--------------+------+-----+--------+
```

The only difference is the MUL value, which is a bit surprising: quoting https://dev.mysql.com/doc/refman/8.0/en/show-columns.html,

> If Key is MUL, the column is the first column of a nonunique index in which multiple occurrences of a given value are permitted within the column.

In other word, this doesn’t really carry any information about the fact that ADDRESS.Habitants is now a foreign key referencing NAME.Id. A way of displaying information about that foreign key is using `SHOW CREATE TABLE`:

```
MariaDB [HW_Address]> SHOW CREATE TABLE ADDRESS;
+-------------+-----------------------------+
| Table       | Create Table                |
+-------------+-----------------------------+
| ADDRESS     | CREATE TABLE `ADDRESS` (`StreetName` varchar(15) NOT NULL, `Number` int(11) NOT NULL, |
```

3.10 Setting Up Your Work Environment

```
`Habitants` int(11) DEFAULT NULL,
PRIMARY KEY (`StreetName`, `Number`),
KEY `Habitants` (`Habitants`),
CONSTRAINT `ADDRESS_ibfk_1` FOREIGN KEY (`Habitants`) REFERENCES
  `NAME` (`Id`) ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 |
+---------+----------------------+
1 row in set (0.01 sec)
```

Pb 3.2 – Solution to Q. 4 We have.

```
<table>
<thead>
<tr>
<th>Name</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>FName</td>
<td>LName</td>
</tr>
<tr>
<td>Id</td>
<td></td>
</tr>
<tr>
<td>StreetName</td>
<td>Number</td>
</tr>
<tr>
<td>Habitants</td>
<td></td>
</tr>
</tbody>
</table>
```

Pb 3.2 – Solution to Q. 5 To display the information back, we can use

```
SELECT * FROM NAME;
```

We can remark that the Id attributes lost their leading zeros.

Pb 3.2 – Solution to Q. 6 This syntax is better for "bulk insertion", since it allows to write less command, and to focus on the data to insert. However, if an error occurs, then nothing gets inserted.

Pb 3.2 – Solution to Q. 7

```
SELECT Id FROM NAME WHERE FName = 'Samantha';
```

Pb 3.2 – Solution to Q. 8 The command

```
INSERT INTO NAME VALUES ('Maria', 'Kashi', NULL);
```

returns

```
ERROR 1048 (23000): Column 'Id' cannot be null
```

Another way of violating the entity integrity constraint is

```
INSERT INTO NAME VALUES ('Maria', 'Kashi', 80);
```

which returns

```
ERROR 1062 (23000): Duplicate entry '80' for key 'PRIMARY'
```

Pb 3.2 – Solution to Q. 9 The command

```
UPDATE ADDRESS SET Habitants = 340 WHERE Number = 120;
```

returns

```
ERROR 1452 (23000): Cannot add or update a child row: a foreign key constraint fails ("HW_Address"."ADDRESS", CONSTRAINT "ADDRESS_ibfk_1" FOREIGN KEY ("Habitants") REFERENCES "NAME" ("Id"))
```

27https://en.wikipedia.org/wiki/Leading_zero
Pb 3.2 – Solution to Q. 10 The query

```sql
INSERT INTO NAME VALUE ('Hi');
```
returns

ERROR 1136 (21S01): Column count doesn’t match value count at row 1

I was violating an implicit constraint, trying to insert a row with fewer values than there are attributes in the table. Note that

```sql
INSERT INTO ADDRESS VALUES ('Maria', 'Random', 98);
```
is a violation of explicit constraint, which is that the value must match the domain (i.e., datatype) of the attribute where they are inserted. However, MySQL or MariaDB doesn’t return an error, and simply replace 'Random' with 0.

Solution to Problem 3.3 (Duplicate rows in SQL) As a preamble, we create our own table:

```sql
CREATE SCHEMA HW_REPETITION;
USE HW_REPETITION;

CREATE TABLE EXAMPLE(
  X VARCHAR(15),
  Y INT
);
```

Pb 3.3 – Solution to Q. 1 The command is

```sql
INSERT INTO EXAMPLE VALUES('Train', 4);
```
If we execute this command twice, then SQL is OK with it, and insert the same tuple twice.

```sql
SELECT * FROM EXAMPLE;
```
displays

```
+-------+------+
| X     | Y    |
|-------+------|
| Train | 4    |
| Train | 4    |
+-------+------+
```
This is an illustration of the fact that a table in SQL is not a set, as opposed to a relation in the relational model.

Pb 3.3 – Solution to Q. 2 The command

```sql
ALTER TABLE EXAMPLE ADD PRIMARY KEY (X);
```
returns

ERROR 1062 (23000): Duplicate entry 'Train' for key 'PRIMARY'
We tried to declare that \( X \) was a primary key, but SQL disagreed, since two rows have the same value for that attribute.

**Pb 3.3 – Solution to Q. 3** Once the table is empty, \( X \) indeed is a primary key, so SQL stops complaining and let us make it a primary key.

**Pb 3.3 – Solution to Q. 4** We cannot introduce the same value twice:

```sql
INSERT INTO EXAMPLE VALUES('Train', 4);
```

returns, the second time:

```
ERROR 1062 (23000): Duplicate entry 'Train' for key 'PRIMARY'
```

which is, by the way, exactly the same error message as when we tried to add the primary key in the first place!

---

**Solution to Problem 3.4 (Revisiting the PROF table)** For Question 1 and 5, we obtain:

![Database Diagram](image)

For the other questions, refer to this code.

```sql
/* code/sql/LECTURE.sql */

-- Question 2
CREATE TABLE LECTURE(
    Name VARCHAR(25),
    Instructor VARCHAR(25),
    Year YEAR(4),
    Code CHAR(5),
```
3.10 Setting Up Your Work Environment

```sql
PRIMARY KEY(Year, Code),
FOREIGN KEY(Instructor) REFERENCES PROF(Login)
);

INSERT INTO LECTURE VALUES
('Intro to CS', 'caubert', 2017, '1304'),
('Intro to Algebra', 'perdos', 2017, '1405'),
('Intro to Cyber', 'aturing', 2017, '1234');

-- This representation can not handle the following situations:
-- - If multiple instructors teach the same class,
-- - If the lecture is taught more than once a year (either because it is taught in the Fall, Spring and Summer, or if multiple sections are offered at the same time),
-- - If a lecture is cross-listed, then some duplication of information will be needed.

-- Question 3
ALTER TABLE GRADE
ADD COLUMN LectureCode CHAR(5),
ADD COLUMN LectureYear YEAR(4);

DESCRIBE GRADE;

SELECT * FROM GRADE;

ALTER TABLE GRADE
ADD FOREIGN KEY (LectureYear, LectureCode)
REFERENCES LECTURE(Year, Code);

-- The values for LectureCode and LectureYear are set to NULL in all the tuples.

-- Question 4
UPDATE GRADE SET LectureCode = '1304', LectureYear = 2017
WHERE Login = 'jrakesh'
AND Grade = '2.85';

UPDATE GRADE SET LectureCode = '1405', LectureYear = 2017
WHERE Login = 'svlatka'
OR (Login = 'jrakesh' AND Grade = '3.85');

UPDATE GRADE SET LectureCode = '1234', LectureYear = 2017
WHERE Login = 'aalyx'
OR Login = 'cjoella';

-- Question 6
SELECT Login, Grade
FROM GRADE
```
3.10 Setting Up Your Work Environment

WHERE Lecturecode='1304'
AND LectureYear = '2017';

SELECT DISTINCT Instructor
FROM LECTURE
WHERE Year = 2017;

SELECT Name, Grade
FROM STUDENT, GRADE
WHERE GRADE.LectureCode = 1405
AND STUDENT.Login = GRADE.Login;

SELECT Year
FROM LECTURE
WHERE Code = '1234';

SELECT Name
FROM LECTURE
WHERE Year IN
(SELECT Year
FROM LECTURE
WHERE Code = '1234');

SELECT B.name
FROM STUDENT AS A, STUDENT AS B
WHERE A.Name = 'Ava Alyx'
AND A.Registered > B.Registered;

SELECT COUNT(DISTINCT PROF.Name) AS 'Head Teaching This Year'
FROM LECTURE, DEPARTMENT, PROF
WHERE Year = 2017
AND Instructor = Head
AND Head = PROF.Login;

Solution to Problem 3.5 (TRAIN table and more advanced SQL coding) Here is some code, commented.

/* code/sql/TRAIN.sql */

-- Question 1:

CREATE TABLE TRAIN(
  Id VARCHAR(30) PRIMARY KEY, -- This line was changed.
  Model VARCHAR(30),
  ConstructionYear YEAR(4)
);

-- Question 2:
CREATE TABLE CONDUCTOR(
    Id VARCHAR(20),
    Name VARCHAR(20),
    ExperienceLevel VARCHAR(20)
);

ALTER TABLE CONDUCTOR
ADD PRIMARY KEY (Id);

-- Question 3

CREATE TABLE ASSIGNED_TO(
    TrainId VARCHAR(20),
    ConductorId VARCHAR(20),
    Day DATE,
    PRIMARY KEY(TrainId, ConductorId),
    FOREIGN KEY (TrainId) REFERENCES TRAIN(Id), -- This line was changed
    FOREIGN KEY (ConductorId) REFERENCES CONDUCTOR(Id) -- This line was changed
);

-- Question 4:

/*
* We insert more than one tuple, to make
* the SELECT statements that follow easier
* to test and debug.
*/

INSERT INTO TRAIN VALUES ('K-13', 'SurfLiner', 2019), ('K-12', 'Regina', 2015);
INSERT INTO CONDUCTOR VALUES ('GP1029', 'Bill', 'Junior'), ('GP1030', 'Sandrine', 'Junior');
INSERT INTO ASSIGNED_TO VALUES ('K-13', 'GP1029', DATE '2015/12/14'),
('K-12', 'GP1030', '20120909');

-- Question 5:

UPDATE CONDUCTOR SET ExperienceLevel = 'Senior' WHERE Id = 'GP1029';

-- Question 6:

-- 1.
SELECT Id FROM TRAIN;

-- 2.
SELECT Name FROM CONDUCTOR WHERE ExperienceLevel = 'Senior';

-- 3.
3.10 Setting Up Your Work Environment

SELECT ConstructionYear FROM TRAIN WHERE Model='SurfLiner' OR Model='Regina';

SELECT ConductorId FROM ASSIGNED_TO WHERE TrainId = 'K-13' AND Day='2015/12/14';

SELECT Model FROM TRAIN, ASSIGNED_TO WHERE ConductorID = 'GP1029' AND TrainId = TRAIN.ID;

Solution to Problem 3.6 (Insert, Select and Project in the COFFEE database) For question 1, we have:

The answer to the other questions can be read from the following code:

/* code/sql/HW_DB_COFFEE.sql */
/
* Setting up the data
*/

--DROP SCHEMA HW_DB_COFFEE;

CREATE SCHEMA HW_DB_COFFEE;
USE HW_DB_COFFEE;

CREATE TABLE COFFEE (Ref VARCHAR(30) PRIMARY KEY,
Origin VARCHAR(30),
TypeOfRoast VARCHAR(30),
PricePerPound DOUBLE
);

CREATE TABLE CUSTOMER (CardNo VARCHAR(30) PRIMARY KEY,
Name VARCHAR(30),
Email VARCHAR(30),
FavCoffee VARCHAR(30))

CREATE TABLE SUPPLY (Coffee VARCHAR(30) PRIMARY KEY, Provider VARCHAR(30))

CREATE TABLE PROVIDER (Name VARCHAR(30), Email VARCHAR(30))
```
3.10 Setting Up Your Work Environment

FavCoffee VARCHAR(30),
FOREIGN KEY (FavCoffee) REFERENCES COFFEE(Ref) ON UPDATE CASCADE
  ON DELETE CASCADE
);
CREATE TABLE PROVIDER(
  Name VARCHAR(30) PRIMARY KEY,
  Email VARCHAR(30)
);
CREATE TABLE SUPPLY(
  Provider VARCHAR(30),
  Coffee VARCHAR(30),
  PRIMARY KEY (Provider, Coffee),
  FOREIGN KEY (Provider) REFERENCES PROVIDER(Name) ON UPDATE CASCADE
  ON DELETE CASCADE,
  FOREIGN KEY (Coffee) REFERENCES COFFEE(Ref) ON UPDATE CASCADE ON
  DELETE CASCADE
);
INSERT INTO COFFEE VALUES
(001, 'Brazil', 'Light', 8.9),
(121, 'Bolivia', 'Dark', 7.5),
(311, 'Brazil', 'Medium', 9.0),
(221, 'Sumatra', 'Dark', 10.25);
INSERT INTO CUSTOMER VALUES
(005, 'Bob Hill', NULL, 001),
(002, 'Ana Swamp', 'swampa@nca.edu', 311),
(003, 'Mary Sea', 'brig@gsu.gov', 121),
(004, 'Pat Mount', 'pmount@fai.fr', 121);
INSERT INTO PROVIDER VALUES
('Coffee Unl.', 'bob@cofunl.com'),
('Coffee Exp.', 'pat@coffeeex.dk'),
('Johns & Co.', NULL);
INSERT INTO SUPPLY VALUES
('Coffee Unl.', 001),
('Coffee Unl.', 121),
('Coffee Exp.', 311),
('Johns & Co.', 221);

-- Question 2:

START TRANSACTION;

INSERT INTO CUSTOMER VALUES(005, Bob Hill, NULL, 001);
INSERT INTO COFFEE VALUES(002, "Peru", "Decaf", 3.00);
INSERT INTO PROVIDER VALUES(NULL, "contact@localcof.com"); -- ERROR
  1048 (23000): Column 'Name' cannot be null
INSERT INTO SUPPLY VALUES("Johns & Co.", 121);
```
3.10 Setting Up Your Work Environment

```sql
INSERT INTO SUPPLY VALUES ("Coffee Unl.", 311, 221); -- ERROR 1136
  (21S01): Column count doesn't match value count at row 1

-- COMMIT;
-- Rest the changes:
ROLLBACK;

-- Question 3:

START TRANSACTION;
UPDATE CUSTOMER SET FavCoffee = '001' WHERE CardNo = '001'; -- Rows
  matched: 1 Changed: 1 Warnings: 0
SELECT * FROM CUSTOMER;
ROLLBACK;

START TRANSACTION;
UPDATE COFFEE SET TypeOfRoast = 'Decaf' WHERE Origin = 'Brazil'; --
  Rows matched: 2 Changed: 2 Warnings: 0
SELECT * FROM COFFEE;
ROLLBACK;

START TRANSACTION;
UPDATE PROVIDER SET Name = 'Coffee Unlimited' WHERE Name = 'Coffee
  Unl.'; -- Rows matched: 1 Changed: 1 Warnings: 0
SELECT * FROM PROVIDER;
SELECT * FROM SUPPLY;
ROLLBACK;

START TRANSACTION;
UPDATE COFFEE SET PricePerPound = 10.00 WHERE PricePerPound > 10.00;
  -- Rows matched: 1 Changed: 1 Warnings: 0
SELECT * FROM COFFEE;
ROLLBACK;

-- Question 4:

START TRANSACTION;
DELETE FROM CUSTOMER WHERE Name LIKE '%S%'; -- Query OK, 2 rows
  affected (0.01 sec)
SELECT * FROM CUSTOMER;
ROLLBACK;

START TRANSACTION;
DELETE FROM COFFEE WHERE Ref = '001'; -- Query OK, 1 row affected (0.00 sec)
SELECT * FROM COFFEE;
SELECT * FROM SUPPLY;
ROLLBACK;

START TRANSACTION;
```
3.10 Setting Up Your Work Environment

DELETE FROM SUPPLY WHERE Provider = 'Coffee Unl.' AND Coffee = '001';
→ Query OK, 1 row affected (0.00 sec)
SELECT* FROM SUPPLY;
ROLLBACK;

START TRANSACTION;
DELETE FROM PROVIDER WHERE Name = 'Johns & Co.'; -- Query OK, 1 row
→ affected (0.00 sec)
SELECT * FROM PROVIDER;
SELECT * FROM SUPPLY;
ROLLBACK;

-- Question 5:

-- 1.
SELECT Origin FROM COFFEE WHERE TypeOfRoast = 'Dark';

-- 2.
SELECT FavCoffee FROM CUSTOMER WHERE Name LIKE 'Bob%';

-- 3.
SELECT Name FROM PROVIDER WHERE Email IS NULL;

-- 4.
SELECT COUNT(*) FROM SUPPLY WHERE Provider = 'Johns & Co.';

-- 5.
SELECT Provider FROM COFFEE, SUPPLY WHERE TypeOfRoast = 'Dark' AND
→ Coffee = Ref;

Solution to Problem 3.8 (Improving a role-playing game) The following solves all the
tofern issues. As quests only rarely provides a special item, we added a relation to
avoid having a Special-item in the QUEST table that would be too often NULL.
Solution to Problem 3.9 (A simple database for books)  Pb 3.9 – Solution to Q. 1

Here are possible ways of getting the required information:

- The title of all the books:
  
  \[ \text{SELECT Title FROM BOOK;} \]

- The (distinct) name of the publishers.
  
  \[ \text{SELECT DISTINCT Name FROM PUBLISHER;} \]

- The title and publication date of the books published since January 31, 2012.
  
  \[ \text{SELECT Title, Published FROM BOOK WHERE Published > DATE'20120131'}; \]

- The first and last names of the authors published by Gallimard (no matter the city).
  
  \[ \text{SELECT FName, LName FROM AUTHOR, BOOK WHERE PublisherName = "Gallimard" AND Author = Id;} \]

- The first and last names of the authors who were not published by an editor in New-York.
  
  \[ \text{SELECT FName, LName FROM AUTHOR, BOOK WHERE NOT PublisherCity= "New-York" AND Author = Id;} \]

- The id of the authors who published a book whose name starts with “Where”.
  
  \[ \text{SELECT Author FROM BOOK WHERE Title LIKE 'Where%';} \]

- The total number of pages in the database.
SELECT SUM(Pages) FROM BOOK;

- The number of pages in the longest book written by the author whose last name is "Wolve".

SELECT MAX(PAGES) FROM BOOK, AUTHOR
WHERE LName = "Wolve"
    AND Author = Id;

- The title of the books published in the XIXth century.

SELECT Title FROM BOOK
WHERE Published >= DATE'18010101'
    AND Published <= DATE'19001231';

Pb 3.9 – Solution to Q. 2 We can use the following command:

UPDATE BOOK SET Title = "BANNED"
WHERE Author = 3;

But, as the pair (title, publication date) is the primary key in the BOOK table, if the author whose id is 3 has published more than one book at a particular date, then our update will be rejected, as applying it would result in violating the entity integrity constraint.

Pb 3.9 – Solution to Q. 3 To delete the required rows, we can use:

DELETE FROM BOOK WHERE Author = 3;
DELETE FROM AUTHOR WHERE Id = 3;

Note that trying to delete the rows in the AUTHOR table before deleting the rows in the BOOK table could cause a referential integrity violation, since some of the books would be "authorless".

Pb 3.9 – Solution to Q. 4 We could design that table as follows:

CREATE TABLE AWARD(
    Name VARCHAR(30),
    Year DATE,
    BookTitle VARCHAR(30),
    BookPubDate DATE,
    FOREIGN KEY (BookTitle, BookPubDate)
        REFERENCES BOOK(Title, Published),
    PRIMARY KEY (Name, Year)
);

Note that there is no need to store the name of the author in that relation: this information can be recovered by looking in the BOOK table for the name of the author of the awarded book.

Pb 3.9 – Solution to Q. 5 Drawing added soon.

Pb 3.9 – Solution to Q. 6 Two of the flaws that come to mind are:

- The choice of the primary key for the BOOK relation: two books with the same title cannot be published on the same day, and that is a serious limitation. Using a primary key like ISBN would be much more appropriate.
- The impossibility to deal with books written by multiple authors or published by multiple publishers. We could address this by having two separate tables,
IS_THE_AUTHOR_OF and PUBLISHED_BY, that “maps” book’s ISBN with author’s or editor’s primary key.
4 Designing a Good Database

Resources

This part of the lecture covers significantly more material than the other, hence we give the details of the references below:

- E.R. models: (Elmasri and Navathe 2010, Ch. 7) or (Elmasri and Navathe 2015, Ch. 3)
- The E.R. to Relational model: (Elmasri and Navathe 2010, Ch. 9.1) or (Elmasri and Navathe 2015, Ch. 9.1)
- Normalization: (Elmasri and Navathe 2010, Ch. 7) or (Elmasri and Navathe 2015, Ch. 3)
- UML: not so much in the textbook, but you can look at (Elmasri and Navathe 2010, Ch. 7.8, 10.3) or (Elmasri and Navathe 2015, Ch. 3.8).

4.1 Interest for High-Level Design

Previous relational models have mistakes and limitations:

- What if a hurricane is over more than one state?
- What if an insurance covers more than one car, more than one driver?
- Changing the code “on the fly”, as we did for the Lecture and Grade tables, is difficult and error-prone.

We could go back and forth between relational models (~ logical level) and SQL implementations (~ physical level), but we will use even more high-level tools (~ conceptual level):

- Entity Relationship Models (ER, static: DB)
- Unified Modelling Diagrams (UML, dynamic: DB + software)
- Enhanced Entity Relationship Models (EER, adds operations to ER)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Conceptual</th>
<th>Logical</th>
<th>Physical</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Main) Audience</td>
<td>Business</td>
<td>Designer</td>
<td>Programmer</td>
</tr>
<tr>
<td>Entity Names</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Entity Relationships</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Attributes</td>
<td>✔ (✓)</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Cardinalities</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Primary Keys</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Foreign Keys</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Column Data types (Domain)</td>
<td>✔ (✓)</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Table Names</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Column Names</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Remember that in relational models, relations were representing entities (Student) and relation-
4.2 Entity-Relationship Model

Data is organized into entities (with attributes), relationships between entities (with attributes as well).

4.2.1 Entities

- Entity = Thing, object, with independent existence.
- Each entity has attributes (properties)

Entity A:

- Name = Clément Aubert
- Address = HCOB, HA, E. 128 ; Invented St., Auguta, GA
- Diploma = Ph.D in CS; BS in Math
- Highest Diploma = Ph.D in CS
- Favorite Class = CSCI 1301
- Favorite Sport = NULL

Some vocabulary:

- Entity = actual thing (individual)
- Entity type = collection of entities with the same attributes
- Entity set (or collection) = collection of all entities of a particular entity type.

4.2.1.1 Attributes

Attributes can be

- Composite (divided in smaller parts) or simple (atomic)
- Single-valued or multi-valued
- Stored vs derived
- Nested!

\{..\} = multi-valued

\(\ldots\) = complex

For instance, one could

- store the name using a composite attribute (First Name, \{Middle Name\}, Last Name),
- store multiple addresses using the "schema" \{Address(Street, Number, Apt, City, State, ZIP)\},
- derive the value of "Highest Diploma" using the value(s) stored in “Diploma”.

4.2.1.2 Key attributes

A key attribute is an attribute whose value is distinct for each entity in the entity set.
• Serve to identify an entity,
• Can be more than one such attribute (and we leave the options open),
• Cannot be multiple attributes: if more than one attribute is needed to make a key attribute, combine them into a composite attribute and make it the key.
• A composite attribute that is a key attribute should not still be a key attribute if we were to remove one of the attribute (similar to the minimality requirement).
• An entity with no key is called a weak entity type: it is an entity that will be identified thanks to its relation to other entities, and thanks to its partial key (we will discuss this later).

4.2.1.3 Drawing entity types

• Entity = squared box (name in upper case)
• Attribute = rounded box connected to square box (name in lower case)

<table>
<thead>
<tr>
<th>If the attribute is ..., then...</th>
</tr>
</thead>
<tbody>
<tr>
<td>composite</td>
</tr>
<tr>
<td>multivalued</td>
</tr>
<tr>
<td>derived</td>
</tr>
<tr>
<td>a key</td>
</tr>
</tbody>
</table>

![Diagram of Entity-Relationship Model]

- Atomic
- Derived
- Key
- Composite
- ENTITY TYPE NAME
- Multi-Valued
- Attribute 1
- ... Attribute n
In the following, we’ll focus on the relationship between the entities more than on the attributes of particular entities, so we’ll sometimes simply draw

leaving the attributes un-specified (but that does not mean that they all have to be atomic) or even just

but that does not mean that the entity type have no attribute!
4.2.2 Relationships

4.2.2.1 Vocabulary

- Relationship = actual relation (or action) between entities ("teaches", "loves", "possesses", etc.).
- Relationship instance = $r_1$ associates $n$ entities $e_1, ..., e_n$ ("Pr. X teaches CSCI YYY", "There is love between Mary and Paul", etc.)
- Relationship set = collection of instances
- Relationship type = abstraction ("Every course belong to one instructor", "Love is a relation between two persons", etc).

$E_1, ..., E_n$ participate in R, $e_1, ..., e_n$ participate in $r_1$, $n$ is the degree.

Note that we can have Entity Set 1 = Entity Set 2, in which case we say the relation is recursive.

Naming convention:

- Use a singular name for entity types.
- Use a verb for relationship.
- Relationship types are drawn in losanges.
- Drawing usually reads right to left, and up to down.
4.2.2.2 Role names and recursive relations

Convenient, and sometimes mandatory, to give role names.

If we want to stress that we are considering only one aspect of an entity (that is, a person is not only an employee, a company is not only an employer, but this aspect is crucial for the "EMPLOYS" relation):

We can also use it to make the “right-side” and the “left-side” of a recursive relationship explicit:
Finally, we will sometimes use “Role Name of Entity 1 : Role Name of Entity 2” as a notation for the relation between them. For instance, we can write “Employer:Employee” to denote the “EMPLOYS” relation, and we will also use this notation when the relationship is between different entities, and write e.g. ”PERSON:POSITION” for the “OCCUPIES” relation.

### 4.2.2.3 Constraints

Two constraints, called ”structural constraints”, applies to relationship types: cardinality ratio and participation constraint. They both concerns the number of relationship instances an entity can participate in (which is different from the cardinality of a relationship type).

#### 4.2.2.3.1 Cardinality ratio

**Maximum** number of relationships instances that an entity can participate it.

For binary relations, can be 1 : 1, N : 1, 1 : N, or M : N. The 1 stands for “at most 1”, and the M, N, and P stand for “possibly more than 1”, or “no maximum”. In E.R. diagram, we don’t count, and don’t make the distinction between “at most 5” and “at most 10”, for instance.

Possible examples include:

<table>
<thead>
<tr>
<th>Relation</th>
<th>Possible Ratio</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MENTOR : MENTEE</td>
<td>1 : N</td>
<td>“A mentor can have multiple mentees, a mentee has at most one mentor.”</td>
</tr>
<tr>
<td>PERSON : SSN</td>
<td>1 : 1</td>
<td>“A person has one SSN, a SSN belongs to one person.”</td>
</tr>
<tr>
<td>COURSE : DEPARTMENT</td>
<td>N : 1</td>
<td>“A course is offered by one department, a department can offer any number of courses”.</td>
</tr>
<tr>
<td>STUDENT : TEAM</td>
<td>M : N</td>
<td>“A student can participate in multiple team, a team can have multiple students.”</td>
</tr>
</tbody>
</table>

We indicate the ratio on the edges:

---

1 An alternative notation, detailed later on, will address this shortcoming.
Note that reflexive relations can have any ratio as well. An example of \( M : N \) recursive relation could be:

This reads "a course must be offered by a department, but a department may or may not offer courses".

**4.2.2.3.2 Participation constraint**

**Minimum** number of relationships instances that an entity can participate in, a.k.a. "minimum cardinality constraint".

The participation can be total (a.k.a. existence dependency, the entity **must** be in that relationship at least once) or partial (the entity may or may not be in that relationship).

Total is drawn with a double line, partial is drawn with a single line:
4.2.2.4 Attributes

Relationships can have attributes too! The typical example is a date attribute.

- TEACHING relation between PROF and CLASS \((N : M)\) could have a “Quarter” attribute.
- MENTORING relation between MENTOR and MENTEE \((1 : N)\) could have a “Since” attribute.
- EMITED_DRIVING_LICENCE between DMV and PERSON \((N : 1)\) could have a “Date” attribute.

We are dealing with moving aspects here: attributes on \(1 : 1\), \(1 : N\), \(N : 1\) relationships can be migrated (to the \(N\) side when there is one, or to either side where there is none).

For instance, imagine that every phone uses exactly (= “at most and at least”) one carrier, that a carrier can provide network to multiple phones, and that the average quality of the network is an attribute in this relationship:

\[
\text{PHONE} \rightarrow \text{USES} \rightarrow \text{CARRIER}
\]

Then each instance of the relation would be of the form (“Phone X”, “Carrier Y”, “9/10”) for some way of ranking the average quality from 0 to 10. Note that, from the fact that the relationship is \(N : 1\), this means that there is only one tuple involving “Phone X”: this means that the average quality could actually be seen as a property of the phone, and hence be migrated as an attribute to the phone side:

\[
\text{PHONE} \rightarrow \text{USES} \rightarrow \text{CARRIER}
\]

Note that we could not migrate the “average phone quality” to the “Carrier” side: imagine if we had the instances (“Phone X”, “Carrier Y”, “9/10”) and (“Phone Z”, “Carrier Y”, “3/10”), then should the attribute of “Carrier Y” be “9/10” or “3/10”: we have no way of deciding based on this model.
As an exercise, you can look at the relationships TEACHING, MENTORING and EMITED_DRIVING_LICENCE that are listed above, and see if the attributes can be migrated or not, and if yes, on which side.

### 4.2.2.5 Relationships of degree higher than two

Of course, relationships can have a degree higher than two. An example of a ternary relation could be:

![Diagram of a ternary relationship]

To determine cardinality ratio, one should fix all but one parameters, and wonder how many values of the remaining parameter can be in that relationship.

For our example, Customer Y and Bank Z could be in relationship with more than 1 account (hence the "N"). On the opposite, Customer Y and Account K would be in relationship with only 1 bank (hence the "1" on the bottom), and Bank Z and Account K would belong to only 1 customer (hence the 1 on the left).

It is sometimes impossible to do without relations with arity greater than 2. For instance, consider the following two diagrams:

![Diagram of a ternary relationship]
You should realize that they convey different information. For instance, you can know for a fact that a person visit a library only if they bought something in it, while the second diagram de-correlate the act of buying with the visit to a library. Similarly, the second diagram could give you a hint that a person that owns a copy of a book Z and visits a library X that sells it could also visit it, but you won’t know that for sure.

An example of recursive ternary relation could be:

An example of relation of degree 4 could be:
4.2.3 Weak entity types

There are actually two sorts of entity types:

- Strong (a.k.a. regular, the ones we studied so far), with a key attribute,
- Weak, without key attribute.

Weak (or child) entity types are identified by identifying / owner type that is related to it, in conjunction with one attribute (the partial key). That relation is called identifying (or supporting) relationship, and weak entities have a total participation constraint. The partial key is an attribute, that, when paired with an entity with which they are in relation through their identifying relationship, allows to identify a particular entity.

Weak entities and identifying relationships have a double border, and partial key have a dotted underline, as follows:
The idea here is that we don’t need to gather data about all the dependent in the world, or in isolation, but are interested in dependent only if they are related to an employee in our database. Just having the name of a dependent is not enough to identify them, but having their name and the SSN of the employee they are related to is enough. The identifying relation always have ratio $1 : M$ or $1 : 1$: a weak entity cannot be related to more than one entity of the owner type, so that $M : N$ ratio are not possible (cf. e.g. https://dba.stackexchange.com/q/17207). If you need to have, for instance, a dependent connected to multiple employees, then that means that your dependent entity should be strong, because it has an existence “of its own”.

You may wonder why we do not represent weak entities simply as (composite, multi-valued) attributes of their owner type. For instance, why would we use

![Diagram 1](image1.png)

instead of

![Diagram 2](image2.png)

The answer depends whenever we need to have the ability to represent our weak entities (here, PET) as being in relationship with other entities (that can themselves be weak!), as follows:
This would be impossible if PET was an attribute of FRIEND! Whenever the pet entity type is involved in other relationships or not should help you in deciding which representation to choose.

- Weak entities types can sometimes be replaced by complex (composite, multi-valued) attributes, unless they are involved in other relationships.
- Owner can itself be weak!
- The degree of the identifying relationship can be more than 2 (cf. e.g., https://stackoverflow.com/q/15393587/).

Another example of weak entity whose owner is weak as well could be:

The idea being that the Health care provider cares about an insuree only if they are covered by them, and that they care about the doula only if they are currently helping one of their insuree.

4.2.4 Alternative Notations

Multiple notations have been used to represent the ratio and constraint on relationship.
Figure 4.1: A Quick Overview of the Notations for ER Diagram (courtesy of wikipedia)
In the following, we introduce two of them: the Min/Max and the Crow’s foot notations.

### 4.2.4.1 Notation With Explicit Maximal (Min/Max Notation)

The two constraints can be written on the same side, and the $N$, $M$, $P$ ratio can be replaced by actual number, providing more information.

For instance,

![Diagram showing the relationship between CAR, CARPOOLING, and PERSON with explicit maxima and minima.]

could be drawn as

![Diagram showing the explicit maxima and minima notation.]

meaning that

- A car can be used to carpool between 1 and 5 persons (and that *it must* be used for at least 1 person),
- A person can be registered for 0, 1, 2 or 3 carpool at the same time.

More generally, we have the following:

\[
\begin{align*}
1 \quad \text{N} & \quad \Rightarrow \quad (0, \text{N}) \quad (0, 1) \\
1 \quad \text{N} & \quad \Rightarrow \quad (1, \text{N}) \quad (0, 1) \\
\text{M} \quad \text{N} & \quad \Rightarrow \quad (1, \text{N}) \quad (1, \text{M})
\end{align*}
\]
4.2.4.2 Crow’s foot notation

One

Many

Exactly one

Zero or one

One or many

Zero or many

4.2.5 Enhanced Entity–Relationship Model

Extended (or Enhanced) E.R. Models (E.E.R.) have additionally:

• Subtype / Subclass: "every professor is an employee". There is a class / subclass relationship (you can proceed by specialization or generalization).
• Category (to represent UNION): an OWNER entity that can be either a PERSON, a BANK, or a COMPANY entity type.

Closer to object-oriented programming.

4.2.6 Reverse Engineering

From relational models to E.R. models (sometimes needed)
4.3 E.R.-to-Relational Models Mapping

4.3.1 Intro

We have to map all of the following:

<table>
<thead>
<tr>
<th>Entity</th>
<th>Strong, Weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes</td>
<td>Composite, Key, Atomic, Multi-valued, Partial Key</td>
</tr>
<tr>
<td>Relationships</td>
<td>Binary (1 : 1, N : 1, 1 : N, N : M), n-ary</td>
</tr>
</tbody>
</table>

Using four tools: Relations, Attributes, Primary Keys, Foreign Keys.

4.3.2 Algorithm

We will use three techniques to represent some of the relationships, the foreign key approach, the merged relations approach and the cross-reference approach. They are detailed and illustrated after the algorithm, which goes as follows:

<table>
<thead>
<tr>
<th>#</th>
<th>is mapped to</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Strong Entity</td>
</tr>
<tr>
<td></td>
<td>Relation with all the simple attributes. Decompose complex (composite) attributes. Pick a key to be the PK, if it is composite, take its elements.</td>
</tr>
</tbody>
</table>
4.3 E.R.-to-Relational Models Mapping

<table>
<thead>
<tr>
<th>#</th>
<th>is mapped to</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Weak Entity</td>
</tr>
<tr>
<td>3</td>
<td>Binary 1 : 1 Relationship Types</td>
</tr>
<tr>
<td>4</td>
<td>Binary 1 : N Relationship Types</td>
</tr>
<tr>
<td>5</td>
<td>Binary M : N Relationship Types</td>
</tr>
<tr>
<td>6</td>
<td>n-ary Relationship Types</td>
</tr>
<tr>
<td>7</td>
<td>Multivalued Attributes</td>
</tr>
</tbody>
</table>

- **Weak Entity**: Relation with all the simple attributes. Decompose complex attributes. Add as a foreign key the primary key of the relation corresponding to the owner entity type, and make it a primary key, in addition to the partial key of the weak entity. If the owner entity type is itself weak, start with it.

- **Binary 1 : 1 Relationship Types**: Foreign Key, Merge Relations or Cross-Reference approach

- **Binary 1 : N Relationship Types**: Foreign Key or Cross-Reference approach

- **Binary M : N Relationship Types**: Cross-Reference approach

- **n-ary Relationship Types**: Cross-Reference approach

- **Multivalued Attributes**: Create a new relation, whose primary key is the foreign key to the relation corresponding to the entity.

1. **Foreign Key Approach**: Chose one of the relationship (preferably with total participation constraint, or on the N side), add a foreign key and all the attributes of the relationship.
2. **Merged Relation Approach**: If both participations are total, just merge them. Primary key = just pick one, and add a NOT NULL constraint on the other.
3. **Cross-Reference or Relationship Relation Approach**: Create a lookup table with an appropriate number of foreign keys, pick some of them (the one on the N side, both if the ratio is M : N, for n-ary it is a bit more complex, cf. example below) as the primary key.

Every time a relationships have attributes, they are mapped to the resulting relation.

Let us look in more details at some of those steps. For strong entities, using steps 1 and 7, the following:

```
<table>
<thead>
<tr>
<th>Serial</th>
<th>COLOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building</td>
<td></td>
</tr>
<tr>
<td>DESK</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>Room</td>
<td></td>
</tr>
</tbody>
</table>
```

would give:
And note that if Serial was a complex attribute, we would just “unfold” it, or decompose it, and make all the resulting attributes the primary key of the relation. If one of the attribute was at the same time multi-valued and composite, as follows:

For relationships, things are a bit more complicated. Consider the following:

Since it is a $1:1$ relationship where one of the side has a partial constraint, we have the choice between two approaches. The foreign key approach would give:
Note that we could also have added the foreign key on the side of ENT.B, referencing the key of ENT.A. But since ENT.A has a total participation constraint, we know that the value of FK will always exist, whereas some entities in ENT.B may not be in relationship with an entity from ENT.A, creating the (nefast) need for \texttt{NULL} values.

For the same diagram, the cross-reference approach would give:

Similarly, note that, in MAPPING, KeyB, or KeyA and KeyB, would also be valid primary keys, but that it makes more sense to have KeyA being the primary key, since we know that ENT.A has a total participation constraint, but ENT.B does not.

If both participation constraints were total, as follows:

Then we could use the merged relations approach, and get:

We picked KeyA to be the primary key for the same reason as before. Note that merging the two entities into one relation also means that you have eventually to do some work on the relations that were referring to them.

Of course, if ENT.A and ENT.B are the same entity (that is, REL is recursive), we would get:
4.3 E.R.-to-Relational Models Mapping

Binary $1 : N$ and binary $M : N$ relationships are dealt with in a similar way, using foreign key or cross-reference approaches. The most difficult part of the mapping is with $n$-ary relationships: we have to use cross-reference approaches, but determining the primary key is not an easy task. Consider the following:

![Diagram](https://via.placeholder.com/150)

The arity constraints here can be rephrased as:

- A member can reserve a particular equipment at multiple time slots (the $N$),
- An equipment can be reserved at a particular time slot by only one member (the 1 on the left),
- A member can reserve only one equipment per time slot (the 1 on the right).

And note that there is no total participation constraint.

To represent the RESERVES relationship, we need to create a relation with attributes referencing the primary key of MEMBER, the primary key of TIME_SLOT, and the primary key of EQUIPMENT. Making them all the primary key does not represent the fact that the same equipment cannot

---

3This development was actually asked at https://dba.stackexchange.com/q/232068.
be booked twice during the same slot, nor that a member can book only one equipment per slot, but allows members to reserve a particular equipment at multiple time slots. To improve this situation, we can either

1. take the foreign key to MEMBER and the foreign key to TIME_SLOT to be the primary key of this relation,
2. or take the foreign key to EQUIPMENT and the foreign key to TIME_SLOT to be the primary key of this relation.

Both solutions enforce only some of the requirement expressed by the E.R. diagram.

### 4.3.3 Outro

<table>
<thead>
<tr>
<th>E.R. Model</th>
<th>Relational Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity type</td>
<td>Entity relation</td>
</tr>
<tr>
<td>1 : 1 or 1 : N relationship type</td>
<td>Foreign key (or relationship relation)</td>
</tr>
<tr>
<td>M : N relationship type</td>
<td>Relationship relation and two foreign keys</td>
</tr>
<tr>
<td>n-ary relationship type</td>
<td>Relationship relation and n foreign keys</td>
</tr>
<tr>
<td>Simple attribute</td>
<td>Attribute</td>
</tr>
<tr>
<td>Composite attribute</td>
<td>Set of simple component attributes</td>
</tr>
<tr>
<td>Multivalued attribute</td>
<td>Relation and foreign key</td>
</tr>
<tr>
<td>Value set</td>
<td>Domain</td>
</tr>
<tr>
<td>Key attribute</td>
<td>Primary key</td>
</tr>
</tbody>
</table>

You can have a look at e.g. [http://holowczak.com/converting-e-r-models-to-relational-models/](http://holowczak.com/converting-e-r-models-to-relational-models/) to get a slightly different explanation of this conversion, and additional pointers.

### 4.4 Guidelines and Normal Form

What makes a good database? At the logical (conceptual) and physical (implementation) levels.

Goals:

1. Information preservation (and avoid loss of information)
2. Minimum redundancy
3. Make queries easy (avoid redundant work, make select / join easy)

For E.R. diagrams, some of the usual techniques\(^4\) are:

- Limit the use of weak entity sets.
- Don’t use an entity when an attribute will do.

### 4.4.1 General rules

#### 4.4.1.1 Semantics

1 relation corresponds to 1 entity or 1 relationship type

---

4.4 Guidelines and Normal Form

4.4.1.2 No anomalies

1. Insertion Anomalies  Having to invent values or to put NULL to insert tuples, especially on a key attribute!

2. Deletion Anomalies  Loosing information inadvertently

3. Modification Anomalies  Updates have to be consistent.

(Bad!) Example:

---------- (Login, Name, AdvisoryName, AdvisorOffice, Major, MajorHead)

----------(Office, PhoneNumber, Building)

1. Advisor without student
2. Delete last student of advisor
3. Advisor change name.

4.4.1.3 Null should be rare

NULL has 3 meanings, wastes space, and makes join / nested projections harder.

Example:

STUDENT(Login, ..., siblingEnrolled)

Transform into “Emergency Contact in University” relation (bonus: allow multiple contacts).

4.4.1.4 Identical Attributes in Different Tables Should Be (Primary, Foreign) Key Pairs

Example with advisorOffice and Office: if we try to write a join to obtain the phone number of a student’s advisor, we will obtain all the phone.

4.4.2 Example

MARKER(Owner, Color, OwnerOffice, Brand, BrandEmail)

TEACHER(Office, Name, Phone)

Corrected to:

MARKER(Owner, Color, Brand)

TEACHER(Office, Name, Phone)

BRAND(Name, Email)
4.4 Guidelines and Normal Form

4.4.3 Functional dependencies

Functional dependencies (FD) is a formal tool used to assess how “good” a database is, a property of the relation schema. Functional dependencies list the constraints between two sets of attributes from the database. For instance, if $X$ and $Y$ are (sets of) attributes, $X \rightarrow Y$ reads “$X$ fixes $Y$”, and implies that the value(s) of $Y$ is fixed by the value(s) of $X$.

4.4.3.1 Using semantics of attributes

“What should be.”

Let us list all the attributes of our previous example:

MARKER.Owner, MARKER.Color, MAKER.Brand, TEACHER.Office, TEACHER.Name, TEACHER.Phone, BRAND.Name, BRAND.Email

Think about their dependencies, and list them:

- TEACHER.Name $\rightarrow$ TEACHER.Office
- BRAND.Name $\rightarrow$ BRAND.Email
- TEACHER.Office $\rightarrow$ TEACHER.Name
- TEACHER.Office $\rightarrow$ TEACHER.Phone
- MAKER.Owner and MARKER.Color $\rightarrow$ MARKER.Brand?

4.4.3.2 Using relation states

“What is.”, can disprove some of the assumptions made previously, but shouldn’t add new dependencies based on it (they may be by chance!).

- Maybe TEACHER.Office $\rightarrow$ TEACHER.Name does not hold, because teachers share offices?
- Maybe TEACHER.Name $\rightarrow$ MARKER.Brand and MARKER.Color seemed to be enforced by the state, but we should not add a functional dependency based on that: there are no “requirement” that a Teacher must always buy the same brand and color, this could simply true be by chance so far and should not be imposed to the teachers.

A particular state cannot enforce a FD, but it can negate one.

Example:

<table>
<thead>
<tr>
<th>Att. 1</th>
<th>Att. 2</th>
<th>Att. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>15</td>
<td>Boston</td>
</tr>
<tr>
<td>Bob</td>
<td>13</td>
<td>Boston</td>
</tr>
<tr>
<td>Jane</td>
<td>12</td>
<td>Augusta</td>
</tr>
<tr>
<td>Emily</td>
<td>12</td>
<td>Augusta</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>May hold</th>
<th>Won’t hold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Att. 2 $\rightarrow$ Att. 3</td>
<td>Att1 $\rightarrow$ Att2</td>
</tr>
<tr>
<td>Att. 3 $\rightarrow$ Att. 2</td>
<td>Att. 3 $\rightarrow$ Att. 2</td>
</tr>
<tr>
<td>Att. 1 $\rightarrow$ Att. 3</td>
<td>Att. 2 $\rightarrow$ Att. 1</td>
</tr>
<tr>
<td>{Att. 1, Att. 2} $\rightarrow$ Att. 3</td>
<td>{Att. 3, Att. 2} $\rightarrow$ Att. 1</td>
</tr>
</tbody>
</table>
4.4 Guidelines and Normal Form

4.4.3.3 Notations

![Diagram](MARKER.Id MARKER.Color MARKER.Brand MARKER.Owner TEACHER.Name TEACHER.Office TEACHER.Phone)

Or, more conveniently:

![Diagram](MARKER (Id Color Brand Owner) TEACHER (Name Office Phone))

If an attribute is a foreign key to another, we will draw an arrow between relations:

![Diagram](TRAIN (Id Model) CONDUCTOR (Name Affected-to))

Note that:

- $X$ and $Y$ are sets, we will write $A$ instead of $\{A\}$, but keep writing $\{A, B\}$ for $\{A, B\}$.
- $\{A_1, ..., A_n\} \rightarrow \{B_1, ..., B_m\}$ means that $A_1$ and ... and $A_n$ fix $B_1$, and that $A_1$ and ... and $A_n$ fix $B_n$, etc.
- $FD_1, FD_2, ..., FD_n$ for the list of functional dependencies, $F$ for all of them.
- $A \rightarrow B$ doesn’t imply nor refute $B \rightarrow A$.
- We won’t write the FD that are implied by (this variation of) Armstrong’s axioms:\footnote{https://en.wikipedia.org/wiki/Armstrong%27s_axioms}:
  - Reflexivity: If $Y$ is a subset of $X$, then $X \rightarrow Y$
  - Augmentation: If $X \rightarrow Y$, then $\{X, Z\} \rightarrow Y$
  - Transitivity: If $X \rightarrow Y$ and $Y \rightarrow Z$, then $X \rightarrow Z$

4.4.3.4 Definitions

Remember superkey (not minimal key), key, candidate key, secondary key? We now have a formal definition.

In one particular relation $R(A_1, ..., A_n)$,

- If $\{A_1, ..., A_n\} \rightarrow Y$ for all attribute $Y$, then $\{A_1, ..., A_n\}$ is a superkey.
- If $\{A_1, ..., A_n\}$ $A_i$ is not a superkey anymore for all $A_i$, then $\{A_1, ..., A_n\}$ is a key.
- We will often discard candidate keys and focus on one primary key.
- If $A_i$ is a member of some candidate key of $R$, it is a prime attribute of $R$. It is a non-prime attribute otherwise.

Given a FD $\{A_1, ..., A_n\} \rightarrow Y$,

- It is a full functional dependency if for all $A_i$, $\{A_1, ..., A_n\} \not\subseteq A_i \rightarrow Y$, does not hold.
• It is a **partial dependency** otherwise.

A FD: \( X \rightarrow Y \) is a **transitive dependency** if there exist a set of attribute \( B \) s.t.

• \( B \neq X, B \neq Y \)
• \( B \) is not a candidate key,
• \( B \) is not a subset of any candidate key,
• \( X \rightarrow B \) and \( B \rightarrow Y \) hold

**Examples on lecture 17’s note to incorporate?**

---

### 4.4.4 Normal forms and keys

First, Second, Third, Fourth, Fifth normal form ("X"NF). Stronger than the Third, there is the Boyce-Codd NF (BCNF)

If you satisfy \( N \), you satisfy \( N - 1 \), \( N - 2 \), etc. The **normal form** of a relation is the highest normal form condition that it meets.

#### 4.4.4.1 First normal form

**Definition**

The domain of all attributes must be atomic (simple, indivisible): exclude multi-valued and composite attributes.

Sometimes, additional requirement that every relation has a primary key. We will take this requirement to be part of the definition of 1NF, but some authors take a relation to be in 1NF if it has at least candidate keys (i.e., multiple possible keys, but no primary key, which makes their definition more general, cf. (Elmasri and Navathe 2015, 14.4.1)). Hence, we will always assume that a primary key is given, and it will be underlined.

**Normalization**

To be written

#### 4.4.4.2 Second normal form

**Definition**

1NF + Every non-prime attribute is fully functionally dependent on the primary key.

**Normalization**

For each attribute \( A \) of the relation whose primary key is \( A_1, \ldots, A_n \):

• Is it prime (i.e., is \( A \in \{ A_1, \ldots, A_n \} \))?
  - Yes → Done.
  - No → Is it partially dependent on the primary key?
    * No, it is fully dependent on the primary key → Done
    * Yes, it depends only of \( \{ A'_1, \ldots, A'_k \} \) → Do the following:
- Create a new relation with \( A \) and \( \{ A'_1, \ldots, A'_k \} \), make \( \{ A'_1, \ldots, A'_k \} \) the primary key, and "import" all the functional dependencies.
- Remove \( A \) from the original relation, and all the functional dependencies that implied it.
- Add a foreign key from \( \{ A'_1, \ldots, A'_k \} \) to their original counterparts in the original relation.

Course (Teacher Code Credit Hours Textbook Difficulty)

becomes

Course (Teacher Code Textbook Difficulty)

CreditHours (Code Credit Hours)

Refinement: note that if more than one attribute depends of the same subset \( \{ A'_1, \ldots, A'_k \} \), we will create two relations: that is useless, we could have created just one. For instance, considering

\[
R( A_1 \ A_2 \ A_3 \ A_4 \ A_5 )
\]

applying the algorithm would give

\[
R( A_1 \ A_2 \ A_5 ) \quad R'( A_2 \ A_3 ) \quad R''( A_2 \ A_4 )
\]

whereas a more subtle algorithm would give

\[
R( A_1 \ A_2 \ A_5 ) \quad R'( A_2 \ A_3 \ A_4 )
\]

Note that in both cases, all the relations are in Second Normal Form, though.

Note also that if our primary key is a singleton, then there is nothing to do, we are in 2NF as soon as we are in 1NF.
4.4 Guidelines and Normal Form

4.4.4.3 Third normal form

4.4.4.3.1 Definition

2NF + no non-prime attribute is transitively dependent on the primary key.

4.4.4.3.2 Normalization

For each attribute $A$ of the relation whose primary key is $A_1, \ldots, A_n$:

- Is it prime (i.e., is $A \in \{A_1, \ldots, A_n\}$)?
  - Yes $\rightarrow$ Done.
  - No $\rightarrow$ Is it transitively dependent on the primary key?
    * No, there is no $\{A'_1, \ldots, A'_k\}$ such that $\{A_1, \ldots, A_n\} \rightarrow \{A'_1, \ldots, A'_k\} \rightarrow A$ and $\{A'_1, \ldots, A'_k\} \not\subseteq \{A_1, \ldots, A_n\}$ and $A \notin \{A'_1, \ldots, A'_n\} \rightarrow$ Done
    * Yes, there is such a $\{A'_1, \ldots, A'_m\} \rightarrow$ Do the following:
      · Create a new relation with $A$ and $\{A'_1, \ldots, A'_k\}$, make $\{A'_1, \ldots, A'_k\}$ the primary key, and import all the functional dependencies,
      · Remove $A$ from the original relation, as well as all the functional dependencies involving it,
      · Add a foreign key from $\{A'_1, \ldots, A'_k\}$ to their original counterparts in the original relation.

4.4.4.4 Examples

We can have a look at another example:

\[
\text{DRIVER (State Driver Licence Num Name Governor)}
\]

Note that $\{\text{State, Driver Licence Num}\}$ would be a valid primary key for this relation, and that adding it would make it a relation in 1NF.

As we can see, the name “Driver” is somehow counter-intuitive, since the relation also carries information about Governors. This relation is actually not in 2NF, because the FD $\{\text{State, Driver Licence Num}\} \rightarrow \text{Governor}$ is not fully functional. A possible way to fix it is to get:

\[
\text{DRIVER (State Driver Licence Num Name) and GOVERNOR (State Governor)}
\]

As you can see, the 2NF helped us in separating properly the entities.

An example of a relation that is in 2NF but not in 3NF could be:
As we can see, all the non-prime attributes are fully functionaly dependent from Login, which is our primary key. But, obviously, one of this dependency is transitive, and breaks the 3NF. A way to fix it is:

As we can see, 3NF also helped us in separating properly the entities, in a slightly different way.

In conclusion, we can observe that every FD $X \rightarrow Y$ s.t. $X$ is a proper subset of the primary key, or a non-prime attribute, is problematic. 2NF is a guarantee that every entity has its own relation, 3NF is a way to avoid data inconsistency.

4.5 Unified Modeling Language Diagrams

4.5.1 Overview

One approach for analysis, design, implementation and deployment of databases and their applications. Databases interact with multiple softwares and users, we need a common language.

Unified Modeling Language\textsuperscript{6} is a standard:

- Generic
- Language-independent
- Platform-independent

Wide, powerful, but also intimidating.

You know UML from object-oriented programming language:

\textsuperscript{6}http://uml.org
That’s a class diagram, there are other types of diagrams, they are not unrelated! For instance, using communication diagrams, deployment diagrams, and state chart diagrams, you can collect the requirements needed to draw a class diagram! They each offer a viewpoint on a software that will help you in making sure the various pieces will fit together: it is a tool commonly used in software engineering, and useful in database design.

4.5.2 Types of diagrams

There are 14 different types of diagrams, divided between two categories: structural and behavioral.

(Source: https://commons.wikimedia.org/wiki/File:UML_diagrams_overview.svg)
4.5.2.1 Structural UML diagrams

They describe structural, or static, relationships between objects, softwares.

- **Class diagram** describes static structures: classes, interfaces, collaborations, dependencies, generalizations, etc. We can represent conceptual database schema with them!
- **Object diagram**, a.k.a. instance diagram, represents the static view of a system at a particular time. You can think of a “freeze” of a program, to be able to observe the value of the variables and the objects (or instances) created.
- **Component diagram** describes the organization and the dependencies among software components (e.g., executables, files, libraries, etc.), to describe how an arbitrary large software system is split into pieces.
- **Deployment diagram** is the description of the physical deployment of artifacts (i.e., software components) on nodes (i.e., hardware). If your program runs on a local computer, fetching data from the Internet, and storing output on a server, you may describe this situation using this sort of diagram.

In this category also exist **Composite structure diagram, Package diagram** and **Profile diagram**.

4.5.2.2 Behavioral UML diagrams

They describe the behavioral, or dynamic, relationship, between components.

- **Use case diagram** describes the interaction between the user and the system. Supposedly, it is the privileged tool to communicate with end-users.
- **State machine diagram**, a.k.a., state chart diagram, describes how a system react to external events. You can picture yourself a complex form of finite state automata diagram.
- **Activity diagram** is a flow of control between activities. You may have seen them already, they are supposedly easy to follow:

![Activity Diagram Example]

Then there is the sub-category of “Interaction diagrams”: 
• **Sequence diagram** describes the interactions between objects over time, the flow of information or messages between objects. It is helpful to grasp the time ordering of the interactions.

• **Communication diagram**, a.k.a., collaboration diagram, describes the interactions between objects as a series of sequenced messages. It is helpful to grasp the structure of the objects, who is interacting with who.

This sub-category also comprise **Timing diagram** and **Interaction overview diagram**.

### 4.5.3 Zoom on classes diagrams

Looking at the “COMPANY conceptual schema in UML class diagram notation”, and comparing it with the "ER schema diagram for the COMPANY database" from the textbook, can help you in writing your own “Rosetta Stone” between ER and UML diagram. Let us introduce some UML terminology for the class diagrams.

<table>
<thead>
<tr>
<th>UML</th>
<th>ER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>Entity Type</td>
</tr>
<tr>
<td>Class Name</td>
<td>Entity Name</td>
</tr>
<tr>
<td>Attributes</td>
<td>Attributes</td>
</tr>
<tr>
<td>Operations (or Method)</td>
<td>Sometimes Derived Attributes</td>
</tr>
<tr>
<td>Association</td>
<td>Relationship Type</td>
</tr>
<tr>
<td>Link</td>
<td>Relationship Instance</td>
</tr>
<tr>
<td>Multiplicities</td>
<td>Structural Constraint</td>
</tr>
</tbody>
</table>

As well as for ER diagram, the domain (or data type) of the attributes is optional. A composite attribute in a ER diagram can be interpreted as a structured domain in a UML diagram (think of a `struct`), and a multi-valued attribute requires to create a new class.

Associations are, to some extent, more expressive than relationship types:

• **As for relationship types**, they can be recursive (or reflexive), and uses role names to clarify the roles of both parties.

• **As for relationship types** they can have attributes: actually, a whole class can be connected to an association.

• **As for relationship types**, they can express a cardinality constraint on the relation between classes. They are written as `min . . . max`, with * for “no maximum”, and the following shorthands: * stands for `0..*` and 1 stands for `1..1`. An association with 1 on one side and * on the other (resp. 1 and 1, * and 1, * and *) is sometimes called “one-to-many” (resp., “one-to-one”, “many-to-one”, “many-to-many”). The notation in partially inverted w.r.t. ER diagrams:
• **As opposed to the relationship types**, they can have a direction, indicating that the user should be able to navigate them only in one direction, or in two (which is the default). This is used for security or privacy purposes.

• **As opposed to the relationship types**, they come in various flavors:
  - You can express **aggregation**, a.k.a. “is part of” relationship, between a whole object and its component (that have their own existence).
  - You can express **composition**, which is the particular case of aggregation where the component doesn’t have an existence of their own.
  - You can express **generalization**, a.k.a. inheritance, that eliminates redundancy and makes a class a specialization of another one.

• **As opposed to the relationship types**, they can be qualified, implying that a class is not connected to the other class as a whole, but to one particular attribute, called the qualifier, or discriminator.

This last feature can be used for weak entities, but not only.
Some of those subtleties depend on your need, and are subjective, but are important tool to design properly a database, and relieving the programmer from the burden of figuring out many details.

**Exercises**

**Exercise 4.1** Name the three high-level models we will be looking at in this class (expanding the acronyms).

**Exercise 4.2** What could be the decomposition of an attribute used to store an email address? When could that be useful?

**Exercise 4.3** Draw the ER diagram for a “COMPUTER” entity that has one multivalued attribute “Operating_System”, a composite attribute “Devices” (decomposed into “Keyboard” and “Mouse”) and an “Id” key attribute.

**Exercise 4.4** Draw the ER diagram for a “CELLPHONE” entity that has a composite attribute “Plan” (decomposed into “Carrier” and “Price”), a “Mobile_identification_number” key attribute, and a multi-valued “App_Installed” attribute.

**Exercise 4.5** Name one difference between a primary key in the relational model, and a key attribute in the ER model.

**Exercise 4.6** Invent an entity type with at least one composite attribute and one atomic attribute, but no multi-valued attribute, identify a possible key attribute, and draw the entity type you obtained using the conventions we used in class.

**Exercise 4.7** What is the degree of a relationship type?

**Exercise 4.8** What is a self-referencing, or recursive, relationship type? Give two examples.
Exercise 4.9 What does it mean for a binary relationship type “Owner” between entity types “Person” and “Computer” to have a cardinality ration $M : N$?

Exercise 4.10 What are the two possible structural constraints on a relationship type?

Exercise 4.11 Under what condition(s) can an attribute of a binary relationship type be migrated to become an attribute of one of the participating entity type?

Exercise 4.12 Suppose a “PRODUCES” relationship with an attribute “Amount” exists between a “PRODUCER” entity type and a “MOVIE” entity type, with ratio $1 : M$. Migrate the “Amount” attribute to one of the entity and draw the resulting diagram.

Exercise 4.13 Express the constraints represented in the following diagram in plain English.

Exercise 4.14 For the following binary relationships, suggest cardinality ratios based on the common-sense meaning of the entity types.

<table>
<thead>
<tr>
<th>Entity 1</th>
<th>Cardinality Ratio</th>
<th>Entity 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>STUDENT</td>
<td>:</td>
<td>MAJOR</td>
</tr>
<tr>
<td>CAR</td>
<td>:</td>
<td>TAG</td>
</tr>
<tr>
<td>INSTRUCTOR</td>
<td>:</td>
<td>LECTURE</td>
</tr>
<tr>
<td>INSTRUCTOR</td>
<td>:</td>
<td>OFFICE</td>
</tr>
<tr>
<td>COMPUTER</td>
<td>:</td>
<td>OPERATING_SYSTEM</td>
</tr>
</tbody>
</table>

Exercise 4.15 Give an example of a binary relationship type of cardinality $1 : N$.

Exercise 4.16 Give an example of a binary relationship type of cardinality $N : 1$, and draw the corresponding diagram (you don’t have to include details on the participating entity types).

Exercise 4.17 Draw an ER diagram with a single entity type, with two stored attributes and one derived attribute. In your answer, it should be clear that the value for the derived attribute can always be obtained from the value(s) for the other attribute(s).

Exercise 4.18 Draw an ER diagram expressing the total participation of an entity type “BURGER” in a binary relation “CONTAINS” between “BURGER” and “INGREDIENT”. What would be the ratio of such a relation?

Exercise 4.19 What is the difference between an entity type and a weak entity type?

Exercise 4.20 What is a partial key?

Exercise 4.21 Why do weak entity type have a total participation constraint?

Exercise 4.22 Convert the following ER diagram into a relational model:
Exercise 4.23  What is insertion anomaly? Give an example.

Exercise 4.24  What is deletion anomaly? Is it a desirable feature?

Exercise 4.25  Why should we avoid attributes whose value will often be NULL? Can the usage of NULL be completely avoided?

Exercise 4.26  Consider the following relation:

\[
\text{PROF(SSN, Name, Department, Bike\_brand)}
\]

Why is it a poor design to have a "Bike\_brand" attribute in such a relation? How should we store this information?

Exercise 4.27  Consider the following relation:

\[
\text{STUDENT(SSN, Name, ... , Sibling\_On\_Campus)}
\]

Why is it a poor design to have a "Sibling\_On\_Campus" attribute in such a relation? How should we store this information?

Exercise 4.28  Consider the following relational database schema:

\[
\text{STUDENT(Login, Name, ..., Major, Major\_Head)}
\]
\[
\text{DEPARTMENT(Code, Name, Major\_Head)}
\]

Assuming that “Major” is a foreign key referencing “DEPARTMENT.Code”, what is the problem with that schema? How could you address it?

Exercise 4.29  Consider the relation \( R(A, B, C, D, E, F) \) and the following functional dependencies:

1. \( F \rightarrow \{D, C\}, D \rightarrow \{B, E\}, \{B, E\} \rightarrow A \)
2. \( \{A, B\} \rightarrow \{C, D\}, \{B, E\} \rightarrow F \)
3. \( A \rightarrow \{C, D\}, E \rightarrow F, D \rightarrow B \)

For each set of functional dependency, give a key for \( R \). We want a key, so it has to be minimal.

**Exercise 4.30** Consider the relation \( R(A, B, C, D, E, F) \) and the following functional dependencies:

\[
A \rightarrow \{D, E\}, D \rightarrow \{B, F\}, \{B, E\} \rightarrow A, \{A, C\} \rightarrow \{B, D, F\}, A \rightarrow F
\]

Answer the following:

1. How many candidate keys is there? List them.
2. How many transitive dependencies can you find? Give them and justify them.

**Exercise 4.31** Consider the relation \( R(A, B, C, D) \) and answer the following:

1. If \( \{A, B\} \) is the only key, is \( \{A, B\} \rightarrow \{C, D\}, \{B, C\} \rightarrow D \) a 2NF? List the nonprime attributes and justify.
2. If \( \{A, B, C\} \) is the only key, is \( A \rightarrow \{B, D\}, \{A, B, C\} \rightarrow D \) a 2NF? List the nonprime attributes and justify.

**Exercise 4.32** Consider the relation \( R(A, B, C, D, E, F) \) with candidate keys \( \{A, B\} \) and \( C \).

Remember that, in all generality, to be a prime attribute, you just need to be part of a possible candidate key. Answer the following:

1. What are the prime attributes in \( R \)?
2. Is \( \{C, D\} \rightarrow E \) a fully functional dependency?
3. Write a set of functional dependencies containing at least one transitive dependency, and justify your answer.

**Exercise 4.33** Consider the relation \( R(A, B, C, D, E) \) and the following functional dependencies:

1. \( C \rightarrow D, \{C, B\} \rightarrow A, A \rightarrow \{B, C, D\}, B \rightarrow E \)
2. \( A \rightarrow \{C, D\}, C \rightarrow B, D \rightarrow E, \{E, C\} \rightarrow A \)
3. \( \{A, B\} \rightarrow D, D \rightarrow \{B, C\}, E \rightarrow C \)

For each one, give one candidate key for \( R \).

**Exercise 4.34** Consider the relation \( R(A, B, C, D, E) \) and answer the following:

1. If \( \{A, B\} \) is the primary key, is \( B \rightarrow E, C \rightarrow D \) a 2NF? List the nonprime attributes and justify.
2. If \( \{A\} \) is the primary key, is \( B \rightarrow C, B \rightarrow D \) a 2NF? List the nonprime attributes and justify.

**Exercise 4.35** Consider the relation \( R(A, B, C, D, E, F) \), and let \( \{B, D\} \) be the primary key, and have additionally the functional dependencies \( \{A, D\} \rightarrow E, C \rightarrow F \). This relation is not in 3NF, can you tell why?

**Exercise 4.36** Consider the relation \( R(A, B, C, D) \) and answer the following:

1. If \( A \) is the only key, is \( A \rightarrow \{B, C, D\}, \{A, B\} \rightarrow C, \{B, C\} \rightarrow D \) a 3NF? List the nonprime attributes and justify.
2. If \( B \) is the only key, is \( B \rightarrow \{A, C, D\}, A \rightarrow \{C, D\}, \{A, C\} \rightarrow D \) a 3NF? List the nonprime attributes and justify.
Exercise 4.37 Consider the relation \( R(A, B, C, D, E) \) and the functional dependencies \( \{A, B\} \rightarrow C, B \rightarrow D, C \rightarrow E \). Answer the following:

1. \( A \) by itself is not a primary key, but what is the only key that contains \( A \)?
2. List the non-prime attributes.
3. This relation is not in 2NF: what transformation can you operate to obtain a 2NF?
4. One of the relation you obtained at the previous step is likely not to be in 3NF. Can you normalize it? If yes, how?

Exercise 4.38 What are the two different categories of U.M.L. diagram?

Exercise 4.39 Can a C++ developer working on Linux and a Java developer working on MacOS use the same class diagram as a basis to write their programs? Justify your answer.

Exercise 4.40 What kind of diagram should we use if we want to ...

1. describe the functional behavior of the system as seen by the user?
2. capture the flow of messages in a software?
3. represent the workflow of actions of an user?

Exercise 4.41 Name two reasons why one would want to use a U.M.L. class diagram over an E.R. diagram to represent a conceptual schema.

Exercise 4.42 Consider the following diagram:

Give the number of attributes for both classes, and suggest two operations for the class that doesn’t have any. Discuss the multiplicities: why did the designer picked those values?

Exercise 4.43 Briefly explain the difference between an aggregation and a composition association.

Exercise 4.44 How is generalization (or inheritance) represented in a U.M.L. class diagram? Why is such a concept useful?

Exercise 4.45 Convert the following E.R. diagram into a U.M.L. class diagram:
Solution to Exercises


Solution 4.2 Name / extension. To have statistics about the extensions, to sort the username by length, etc.

Solution 4.3
Solution 4.4

Solution 4.5 There can be more than one key in the ER model, but it has to be made of a single attribute, whereas a primary key can be made of multiple attributes.

Solution 4.6

Solution 4.7 The number of participating entity types.

Solution 4.8 A relationship type where the same entity type participates more than once. On seats, “is to the left of”, on persons, “is married to”.

Solution 4.9 That a person can own multiple computers, and that a computer can have multiple owners.

Solution 4.10 Cardinality ration and participation constraints.

Solution 4.11 When the cardinality is $1: N$, $1: 1$ or $N: 1$.

Solution 4.12 We could have the following:

Solution 4.13 A key opens only one door, and every key must open at least one door. A door can be opened by multiple key, and some doors may not be opened by keys (think of doors
that cannot be locked).

**Solution 4.14**

<table>
<thead>
<tr>
<th>Entity 1</th>
<th>Cardinality Ratio</th>
<th>Entity 2</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>STUDENT</td>
<td>$N : 1$</td>
<td>MAJOR</td>
<td>“A student has one major, but multiple students can have the same major”</td>
</tr>
<tr>
<td>CAR</td>
<td>$1 : 1$</td>
<td>TAG</td>
<td>“A car has exactly one tag, a tag belongs to one particular car.”</td>
</tr>
<tr>
<td>INSTRUCTOR</td>
<td>$1 : N$</td>
<td>LECTURE</td>
<td>“An instructor can teach multiple lecture, but a lecture is taught by only one person.”</td>
</tr>
<tr>
<td>INSTRUCTOR</td>
<td>$1 : N$</td>
<td>OFFICE</td>
<td>“An instructor can have multiple office, but an office belongs to only one instructor”</td>
</tr>
<tr>
<td>COMPUTER</td>
<td>$M : N$</td>
<td>OS</td>
<td>“A computer can have multiple operating system, the same operating system can be installed on more than one computer.”</td>
</tr>
</tbody>
</table>

Some of those choices, of course, are arguable (typically, almost any combination seems reasonable for the INSTRUCTOR : OFFICE relation).

**Solution 4.15** SUPERVISION, a recursive relationship on EMPLOYEE.

**Solution 4.16**

**Solution 4.17**
Solution 4.18

Solution 4.19 The weak entity type doesn’t have a key attribute, it cannot be distinguished from the other weak entities based on a single attribute, for that we also need to know its relationship to some other entity type.

Solution 4.20 For a weak entity attribute, it is the attribute that can uniquely identify weak entities that are related to the same owner entity.

Solution 4.21 Otherwise, we couldn’t identify entities in it without owner entity.

Solution 4.22 A possible option is:

```
<table>
<thead>
<tr>
<th>NAME</th>
<th>ADDRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSN</td>
<td>DOB</td>
</tr>
<tr>
<td>StaysAt</td>
<td></td>
</tr>
<tr>
<td>Address</td>
<td>Rooms</td>
</tr>
</tbody>
</table>
```

Note that “Stays_At” could also be a separate relation, with two attributes, “Address” and “Person”, linked to respectively PLACE.Address and PERSON.SSN, and both being the primary key of the relation.

Solution 4.23 When you have to invent a primary key or add a lot of NULL value to be able to add a tuple. I want to add a room in my DB, but the only place where rooms are listed are as an attribute on a Instructor table, so I have to “fake” an instructor to add a room.

Solution 4.24 A delete anomaly exists when certain attributes are lost because of the deletion of other attributes. It is not desirable, since it can lead to the loss of information.

Solution 4.25 Because they waste space, they are ambiguous (N/A, or unknown, or not communicated?), and they make queries harder. No, it is necessary sometimes.

Solution 4.26 Because it will be NULL most of the time. In a separate relation, e.g. a “BIKE” relation, with two attributes, “Owner” and “Brand”, “Owner” being a foreign key referencing the SSN attribute of PROF.

Solution 4.27 Because it will be NULL most of the time, and because students could have more than one sibling on campus. In a separate relation, e.g. in a “EMERGENCY_CONTACT” relation, with two attributes, “Student” (referencing the SSN attribute of STUDENT), and “Contact”. If the emergency contacts are not related to the student, or if we want to preserve the fact that one student is a sibling to another, we can create another relation to store that information.

Solution 4.28 Major_Head will give update anomalies. By putting the Head of the department in the DEPARTMENT relation only, i.e., removing it from STUDENT.
Solution 4.29  
1. $F$
2. $\{A, B, E\}$
3. $\{A, E\}$

Solution 4.30  
1. Only one: $\{A, C\}$,
2. $A \rightarrow F$ by $A \rightarrow D, D \rightarrow F$.

Solution 4.31  
1. Yes. $C$ and $D$ are non prime, and they fully depend on $\{A, B\}$.
2. No. $D$ is the only non prime, and it depends only on $A$.

Solution 4.32  
1. $A, B$ and $C$.
2. No, because we can remove $D$,
3. $A \rightarrow D, D \rightarrow E$ and $A \rightarrow E$

Solution 4.33  
1. $\{B, C\}, A$
2. $A, \{C, E\}$,
3. $\{A, D, E\}, \{A, B, E\}$

Solution 4.34  
1. No. $C, D, E$, and $E$ has a partial relation to $B$
2. Yes. Since the primary key is a singleton, it is obvious.

Solution 4.35  
$\{B, D\} \rightarrow C \rightarrow F$ breaks the 3NF.

Solution 4.36  
1. No. $B, C$ and $D$ are non prime, $A \rightarrow \{B, C\} \rightarrow D$ breaks the 3NF.
2. No. $A, B$ and $D$ are non prime, $B \rightarrow \{A, C\} \rightarrow D$ breaks the 3NF.

Solution 4.37  
1. $\{A, B\}$,
2. $C, D, E$,
3. $R_3(A, B, C, E)$ and $R_2(B, D)$
4. $R_1(A, B, C), R_2(C, E)$ and $R_3(B, D)$

Solution 4.38  The two different categories of U.M.L. diagram are behaviour and structure.

Solution 4.39  Yes, U.M.L. diagram is language-independent and platform-independent.

Solution 4.40  
1. Use-case
2. Sequence diagram
3. Activity diagram

Solution 4.41 To use direction for association, to have a common language with someone less knowledgeable of other diagrammatic notations. For the concept of integration.

Solution 4.42 Flight has 5 attributes, Plane has 4. The Plane class could have the operations  
\textit{getLastFlightNumber()} : Integer and \textit{setMaximumSpeed(MPH)} : void.

For the multiplicities: A flight could not have a plane assigned, and a plane could not be assigned to a flight. A plane can be assigned to multiple (or no) flights, but a flight must have at most one plane (and could have none).

Solution 4.43 Aggregation: associated class can have an existence of its own.
Composition association: class doesn’t exist without the association.

Solution 4.44 Because it avoids redundancy.
4.5 Unified Modeling Language Diagrams

Solution 4.45

Problems

Problem 4.1 *(Design for your professor)* Your professor designed the following relational model at some point in his career, to help him organizing his exams and the students grades:

<table>
<thead>
<tr>
<th>Table Name and Attributes</th>
<th>Example of Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXAM(Number, Date, Course)</td>
<td>&lt; 1, '2018-02-14', 'CSCI3410'&gt;</td>
</tr>
<tr>
<td>PROBLEM(Statement, Points, Length, Exam)</td>
<td>&lt; 'Your professor designed...', 10, '00:10:00', 1&gt;</td>
</tr>
<tr>
<td>STUDENT_GRADE(Login, Exam, Grade)</td>
<td>&lt; 'aalyx', 1, 83&gt;</td>
</tr>
</tbody>
</table>

where EXAM.Number, PROBLEM.Statement, STUDENT_GRADE.Login and STUDENT_GRADE.Exam are primary key, and where STUDENT GRADE.Exam and PROBLEM.Exam both refer to EXAM.Number.

The idea was to have

- The EXAM table storing information about exams,
- One entry per problem in the PROBLEM table, and to associate every problem to an exam,
- The grade of one student for one particular exam stored in the STUDENT GRADE table.

Unfortunately, this design turned out to be terrible. Describe at least one common and interesting situation where this model would fail to fulfill its purpose, and propose a way to correct the particular problem you identified.

Problem 4.2 *(Reading the MOVIES database ER schema)* Consider the ER schema for the MOVIES database ((Elmasri and Navathe 2010, Figure 7.24)):
Assume that MOVIES is a populated database. ACTOR is used as a gender-neutral term. Given the constraints shown in the ER schema, respond to the following statements with True or False. Justify each answer.

1. There are no actors in this database that have been in no movies.
2. There might be actors who have acted in more than ten movies.
3. Some actors could have done a lead role in multiple movies.
4. A movie can have only a maximum of two lead actors.
5. Every director have to have been an actor in some movie.
6. No producer has ever been an actor.
7. A producer cannot be an actor in some other movie.
8. There could be movies with more than a dozen actors.
9. Producers can be directors as well.
10. A movie can have one director and one producer.
11. A movie can have one director and several producers.
12. There could be some actors who have done a lead role, directed a movie, and produced some movie.
13. It is impossible for a director to play in the movie (s)he directed.

Problem 4.3 (ER diagram for car insurance) Draw the ER diagram for the following situation: A car-insurance company wants to have a database of accidents. An accident involves cars, drivers, and it has several aspects: the moment and place where it took place, the amount of damages, and a (unique) report number. A car has a license, a model, a year, and an owner. A driver has an id, an age, a name, and an address.

One of the interesting choice is: should “accident” be an entity type or a relationship type?
**Problem 4.4** *(Reverse engineering by hand)* Look at the following relational model, and “reverse-engineer” it to obtain an E.R. diagram:

![ER Diagram](image)

**Problem 4.5** *(Discovering MySQL Workbench)* In this problem, we will install and explore the basic functionalities of MySQL Workbench, which is a cross-platform, open-source, and free graphical interface for database design.

1. Install MySQL Workbench if needed. Maybe you already included it in the packages to install when you installed MySQL (cf. the instructions to install MySQL on Windows): try to find if this is the case before trying to install it. Otherwise, use your package manager, or download the binaries from [https://dev.mysql.com/downloads/workbench/](https://dev.mysql.com/downloads/workbench/). The installation should be straightforward for all operating system.

2. Once installed, execute the software. The instructions below were tested for the 6.3.8 version on Debian, and 8.0.15 version on Windows. The trouble with GUI-software is that the menus may differ slightly with what you see, but the core tools we will be using should still be there, and under a similar name, if not the same.

3. Under the panel “MySQL Connections”, you should see your local installation listed as “Local instance 3306”. Click on the top-right corner of that box, and then on “Edit Connections”. Alternatively, click on “Database”, on “Manage Connections”, and then on “Local instance 3306”.

4. Check that all the parameters are correct. Normally, you only have to change the name of the user to “testuser”, and leave the rest as it is. Click on “Test the connection”, and enter your password (which should be “password”) when prompted. If you receive a warning about “Incompatible/nonstandard server version or connection protocol detected”, click on “Continue anyway”.

5. Now, click on the box “Local instance 3306”, and enter your password. A new tab appears, you can see the list of schemas in the bottom part of the left panel.

6. Click on “Database”, and then on “Reverse Engineering” (or hit ctr+z), click on “next”, enter your password, and click on “next”. You should see the list of the schemas stored in your database. Select one (any one, we are just exploring the functionalities at that point, you can pick for instance HW_DB_COFFEE from Problem 4.5 (Discovering...
4.5 Unified Modeling Language Diagrams

MySQL Workbench), click on “next”, and then click on “execute”, “next”, and “close”.

7. You’re back on the previous view, but you should now see “E.E.R. diagram” on the top of the middle panel. Click on “E.E.R. diagram” twice, scroll down if needed, and you should see the E.E.R. diagram.

8. This diagram isn’t exactly an E.R. diagram, and it is not a U.M.L. diagram either: it is an E.E.R. diagram, that uses Crow’s foot notation. Make sure you understand it.

9. Try to modify the E.E.R. diagram. Make some relations mandatory, change their name, add an attribute, change the name of another, insert a couple of elements in an entity, add a row in a table, etc. Make sure you understand the meaning of the lines between the entities.

10. Once you’re done, try to “Forward Engineer” by hitting “Ctrl” + “G”. Click on “next” twice, enter your password, click on “next” once more, and you should see the SQL code needed to produce the table you just designed using the graphical tool.

Problem 4.6 (ER-to-Relation mapping for car insurance)

Apply the ER-to-Relation mapping to your ER diagram from Problem 4.3 (ER diagram for car insurance).

Problem 4.7 (From E.R. diagram to Relational model – BIKE) Consider the following E.R. diagram:

Based on this diagram, answer the following: “Is it true that ...”

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>... a customer cannot drop two bikes at the exact same time and date?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>... two different customers cannot drop two different bikes at the exact same time and date?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>... an employee cannot repair two bikes at the same time?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
"... a customer can be assigned to more than one employee?"
"... a customer can have a bike repaired by an employee that is not assigned to them?"
"... a bike can be in the database without having been dropped by a customer?"
"... an employee can be asked to repair a bike without having that type of bike as one of their specialty?"

Then, convert that E.R. diagram to the relational model. Try to make as few assumptions as possible.

**Problem 4.8 (ER-to-Relation mapping for Country)** Consider the following E.R. schema:

\[
\begin{align*}
\text{COUNTRY} & \rightarrow \text{Name, Population} \\
\text{SPEAKS} & \rightarrow \text{M} \\
\text{LANGUAGE} & \rightarrow \text{N} \\
\text{NAME} & \rightarrow \text{M} \\
\end{align*}
\]

where

- "W_IN" stands for "WRITTEN_IN", and
- "B_W_F" stands for "BORROWS_WORDS_FROM".

For this relationship, on the left-hand side is the language that borrows a word, and on the right-hand side is the language that provides the loanword.

Map that E.R. diagram to a relational database schema.

**Problem 4.9 (From business statements to E.R. diagram – UNIVERSITY)** Consider the following requirements for a UNIVERSITY database, used to keep track of students' transcripts.
1. The university keeps track of each student’s name, student number, class (freshman, sophomore, ..., graduate), major department, minor department (if any), and degree program (B.A., B.S., ..., Ph.D.). Student number has unique values for each student.
2. Each department is described by a name and has a (unique) department code.
3. Each course has a course name, a course number, credit hours, and is offered by at least one department. The value of course number is unique for each course. A course has at least one section.
4. Each section of a course has an instructor, a semester, a year, and a section number. The section number distinguishes different sections of the same course that are taught during the same semester/year; its values are 1, 2, 3, ..., up to the number of sections taught during each semester. Students can enroll in sections and receive a letter grade, and grade point (0, 1, 2, 3, 4 for F, D, C, B, A, respectively).

Draw an E.R. diagram for that schema. Specify key attributes of each entity type and structural constraints on each relationship type. Note any unspecified requirements, and make appropriate assumptions to make the specification complete.

**Problem 4.10 (Normal form of a CAR_SALE relation)** Consider the following relation, and its functional dependencies:

**CAR_SALE(Car_no, Date_sold, Salesman_no, Commission, Discount_amt)**

<table>
<thead>
<tr>
<th>Car_no, Salesman_no</th>
<th>Date_sold, Commission, Discount_amt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date_sold</td>
<td>Discount_amt</td>
</tr>
<tr>
<td>Salesman_no</td>
<td>Commission</td>
</tr>
</tbody>
</table>

and let \{Car_no, Salesman_no\} be the primary key of this relation.

Based on the given primary key, is this relation in 1NF, 2NF, or 3NF? Why or why not? Normalize it to its third normal form.

**Problem 4.11 (Normalizing the FLIGHT relation)** Consider the following relation:

**FLIGHT(From, To, Airline, Flight, DateHour, HeadQuarter, Pilot, TZDifference)**

A tuple in the FLIGHT relation contains information about an airplane flight: the airports of departure and arrival, the airline carrier, the number of the flight, its time of departure, the headquarter of the company chartering the flight, the name of the pilot(s), and the time zone difference between the departure and arrival airports.

The “Pilot” attribute is multi-valued (so that between 1 and 4 pilot’s names can be stored in it). Given an airline and a flight number, one can determine the departure and arrival airports, as well as the date and hour and the pilot(s). Given the airline carrier, one can determine the headquarter. Finally, given the departure and arrival airports, one can determine their time zone difference.

Normalize the “FLIGHT” relation to its third normal form. You can indicate your steps, justify your reasoning, and indicate the foreign keys if you want to, but don’t have to.
**Problem 4.12 (From business statement to dependencies)** This problem asks you to convert business statements into dependencies. Consider the following relation:

\[
\text{BIKE}(\text{Serial_no, Manufacturer, Model, Batch, Wheel_size, Retailer})
\]

Each tuple in the relation BIKE contains information about a bike with a serial number, made by a manufacturer, with a particular model number, released in a certain batch, which has a certain wheel size, and is sold by a certain retailer.

- Write each of the following dependencies as a functional dependency (the first one is given as an example):
  1. A retailer cannot have two bikes of the same model from different batches. solution: \{Retailer, Model\} \rightarrow \text{Batch}
  2. The manufacturer and serial number uniquely identifies the bike and where it is sold.
  3. A model number is registered by a manufacturer and therefore cannot be used by another manufacturer.
  4. All bikes in a particular batch are of the same model.
  5. All bikes of a certain model have the same wheel size.

- Based on those statements, what could be a key for this relation?

- Assuming all those functional dependencies hold, and taking the primary key you identified at the previous step, what is the degree of normality of this relation? Justify your answer.

**Problem 4.13 (Normalization)** Consider the relations \(R\) and \(T\) below, and their functional dependencies (on top of the one induced by the primary keys):

\[
R(\text{EventId, Email, Time, Date, Location, Status})
\]

\[
T(\text{Invno, Subtotal, Tax, Total, Email, Lname, Fname, Phone})
\]

\[
\{\text{EventId, Email}\} \rightarrow \text{Status}
\]

\[
\text{EventId} \rightarrow \{\text{Time, Date, Location}\}
\]

\[
\text{Invno} \rightarrow \{\text{Subtotal, Tax, Total, Email}\}
\]

\[
\text{Email} \rightarrow \{\text{Fname, Lname, Phone}\}
\]

Normalize the relations to 2NF and 3NF. Show all relations at each stage (2NF and 3NF) of the normalization process.

**Problem 4.14 (Normal form of the BOOK relation)** Consider the following relation for published books:

\[
\text{BOOK}(\text{Book_title, Book_type, Author_name, List_price, Author_affil, Publisher})
\]

Suppose we have the following dependencies:

\[
\begin{align*}
\text{Book_title} & \rightarrow \{\text{Publisher, Book_type}\} \\
\text{Book_type} & \rightarrow \text{List_price}
\end{align*}
\]
Problem 4.15 (Normal form of the CONTACT relation) Consider the relation

\[ \text{CONTACT(Phone, Call\_center, Email, Zip, Brand, Website)} \]

and the following functional dependencies:

\[
\begin{align*}
\{\text{Zip, Brand}\} &\rightarrow \{\text{Phone}\} \\
\{\text{Brand}\} &\rightarrow \{\text{Email}\} \\
\{\text{Brand}\} &\rightarrow \{\text{Website}\} \\
\{\text{Phone}\} &\rightarrow \{\text{Call\_center}\}
\end{align*}
\]

Assume that \{Zip, Brand\} is the primary key. Normalize this relation to the second normal form, and then to the third normal form. Give the relations, their primary keys, and functional dependencies for both steps.

Problem 4.16 (CONSULTATION relation: justification, primary key and normal form) Consider the relation

\[ \text{CONSULTATION(Doctor\_no, Patient\_no, Date, Diagnosis, Treatment, Charge, Insurance)} \]

with the following functional dependencies:

\[
\begin{align*}
\{\text{Doctor\_no, Patient\_no, Date}\} &\rightarrow \{\text{Diagnosis}\} \\
\{\text{Doctor\_no, Patient\_no, Date}\} &\rightarrow \{\text{Treatment}\} \\
\{\text{Treatment, Insurance}\} &\rightarrow \{\text{Charge}\} \\
\{\text{Patient\_no}\} &\rightarrow \{\text{Insurance}\}
\end{align*}
\]

1. The designer decided not to add the functional dependency \{Diagnosis\} \rightarrow \{Treatment\}. Explain what could be the designer’s justification, at the level of the mini-world.
2. Identify a primary key for this relation.
3. What is the degree of normalization of this relation? Normalize it to the third normal form if necessary.

Problem 4.17 (From E.R. to relational schema and UML class diagram – CAR_INFO) Consider the following E.R. schema for the CAR_INFO database:
Note that a car can have at most one driver, \( N \) passengers, \( N \) insurances, and that car insurances exist only if they are “tied up” to a car (i.e., they are weak entities, and their identifying relationship is called “Insured”).

1. Find the key attribute for “Car”, and the partial key for “Car Insurance”. If you cannot think of any, add a dummy attribute and make it be the key.
2. Convert that E.R. diagram to a relational database schema.
3. Convert the E.R. diagram to a U.M.L. class diagram. Comparing Figure 7.16 with Figure 7.2 from your textbook should guide you.

**Problem 4.18 (Using MySQL Workbench’s reverse engineering)** This problem requires you to have successfully completed Pb 4.5 and Pb 4.20.

Using the relational database schema you obtained in Pb 4.20, write the SQL implementation of that database. Then, using MySQL Workbench, use the “Reverse Engineering” function to obtain a E.E.R. diagram of your database, and compare it with the U.M.L. diagram from Pb 4.20. Apart from the difference inherent to the nature of the diagram (i.e., U.M.L. Vs E.E.R.), how are they the same? How do they differ? Is the automated tool as efficient and accurate as you are?

**Problem 4.19 (From business statements to dependencies – KEYBOARD)** This exercise asks you to convert business statements into dependencies. Consider the following relation:

\[
\text{KEYBOARD(Manufacturer, Model, Layout, Retail\_Store, Price)}
\]

A tuple in the KEYBOARD relation contains information about a computer keyboard: its
manufacturer, its model, its layout (AZERTY, QWERTY, etc.), the place where it is sold, and its price.

1. Write each of the following business statement as a functional dependency:
   a) A model has a fixed layout.
   b) A retail store cannot have two different models produced by the same manufacturer.

2. Based on those statements, what could be a key for this relation?

3. Assuming all those functional dependencies hold, and taking the primary key you identified at the previous step, what is the degree of normality of this relation? Justify your answer.

Problem 4.20 (From UML to relational model – DRIVER) Consider the UML diagram below, and convert it to the relational model. Don’t forget to indicate primary and foreign keys.

Solution to Selected Problems

Solution to Problem 4.2 (Reading the MOVIES database ER schema)

1. true
2. true
3. true
4. true
5. false
6. false
7. false
8. true
9. true
10. true
11. true
12. true
13. false
Solution to Problem 4.3 (ER diagram for car insurance)

OR

Solution to Problem 4.7 (From E.R. diagram to Relational model – BIKE) “Is it true that ...”

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;... a customer cannot drop two bikes at the exact same time and date?</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>&quot;... two different customers cannot drop two different bikes at the exact same time and date?</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>&quot;... an employee cannot repair two bikes at the same time?</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>&quot;... a customer can be assigned to more than one employee?</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>&quot;... a customer can have a bike repaired by an employee that is not assigned to them?</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
For the $1:M$ relationships that are not identifying, we can choose between the foreign key and the cross-reference approaches. If we use the former, we obtain:

We could also have used a combination of both!

**Solution to Problem 4.10 (Normal form of a CAR_SALE relation)** The CAR_SALE relation is in 1st normal form, since it has a primary key, and by assuming that all the attributes are atomic. This relation is not in 2nd Normal Form: since Date_sold $\rightarrow$ Discount_amount and Salesman_no $\rightarrow$ Commission, then some attributes (namely Discount_amount and Commission) are not fully functional dependent on the primary key. Hence, this relation cannot be in 3rd normal form either.

To normalize,

2NF:

<table>
<thead>
<tr>
<th>Relations</th>
<th>Functional Dependencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car_Sale1(Car_no, Date_sold, Discount_amt)</td>
<td>Car_no $\rightarrow$ {Date_Sold, Discount_amt} and Date_Sold $\rightarrow$ Discount_amt</td>
</tr>
<tr>
<td>Car_Sale2(Car_no, Salesman_no)</td>
<td>Car_no $\rightarrow$ Salesman_no</td>
</tr>
<tr>
<td>Car_Sale3(Salesman_no, Commission)</td>
<td>Salesman_no $\rightarrow$ Commission</td>
</tr>
</tbody>
</table>

3NF:

<table>
<thead>
<tr>
<th>Relations</th>
<th>Functional Dependencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car_Sale1-1(Car_no, Date_sold)</td>
<td>Car_no $\rightarrow$ Date_Sold</td>
</tr>
<tr>
<td>Car_Sale1-2(Date_sold, Discount_amt)</td>
<td>Date_Sold $\rightarrow$ Discount_amt</td>
</tr>
<tr>
<td>Car_Sale2(Car_no, Salesman_no)</td>
<td>Car_no $\rightarrow$ Salesman_no</td>
</tr>
</tbody>
</table>
 Relations | Functional Dependencies
---|---
Car_Sale3(Salesman_no,Commission) | Salesman_no → Commission

**Solution to Problem 4.12 (From business statement to dependencies)**

1. \{ Manufacturer, Serial_no \} → \{ Model, Batch, Wheel_size, Retailer \}
2. Model → Manufacturer
3. Batch → Model
4. \{ Model, Manufacturer \} → Wheel_size

- \{ Manufacturer, Serial_no \}
- If every attribute is atomic, it is in second normal form. \{ Manufacturer, Serial_no \} → Batch → Model breaks the 3NF.

**Solution to Problem 4.16 (CONSULTATION relation: justification, primary key and normal form)**

1. The treatment for a particular disease can vary with the patient (for instance, his age can be a crucial parameter).
2. \{ Doctor_no, Patient_no, Date \} is a primary key for this relation.
3. Because there are no partial dependencies, the given relation is in 2NF already. This however is not 3NF because the Charge is a nonkey attribute that is determined by another nonkey attribute, Treatment. We must decompose further:

    CONSULTATION (Doctor_no, Patient_no, Date, Diagnosis, Treatment)
    PRICE_LISTING (Treatment, Charge)

**Solution to Problem 4.14 (Normal form of the BOOK relation)**

- \{ Book Title, Author Name \}
- If an attribute is composite or multi-valued.
- Because of \{ Book_title \} → \{ Publisher, Book_type \}. We can normalize it as (Book Title, Publisher, Book Type, List Price), (Author Name, Author Affiliation), (Author Name, Book Title).
- Because of \{ Book title \} → \{ Book_type \} → \{ List_price \} \{ Book Title, Publisher, Book Type \} and (Book Type, List Price), (Author Name, Author Affiliation), (Author Name, Book Title).

**Solution to Problem 4.17 (From E.R. to relational schema and UML class diagram – CAR_INFO)**

For “Car”, we need to create an attribute, like “vin”. For “Car Insurance”, “Policy Number” is perfect.
4.5 Unified Modeling Language Diagrams

Note that, during the conversion, we had to make "Insured Car" part of the primary key of CAR INSURANCE.

Solution to Problem 4.18 (Using MySQL Workbench’s reverse engineering) We give first the code, then the drawing.

```sql
/* code/sql/PERSON.sql */
-- DROP SCHEMA HW_Queries;

CREATE SCHEMA HW_6;

USE HW_6;
```
CREATE TABLE PERSON(
    id VARCHAR(25) PRIMARY KEY,
    name VARCHAR(25),
    street VARCHAR(25),
    city VARCHAR(25),
    seat VARCHAR(25),
    position VARCHAR(25)
);

CREATE TABLE CAR(
    vin VARCHAR(25) PRIMARY KEY,
    make VARCHAR(25),
    model VARCHAR(25),
    year DATE,
    driver VARCHAR(25),
    FOREIGN KEY (driver) REFERENCES PERSON(id)
    ON UPDATE CASCADE
);

ALTER TABLE person ADD FOREIGN KEY (seat) REFERENCES car(vin);

CREATE TABLE CAR_INSURANCE(
    policy_number VARCHAR(25) PRIMARY KEY,
    company_name VARCHAR(25),
    insured_car VARCHAR(25),
    FOREIGN KEY (insured_car) REFERENCES CAR(vin)
);

CREATE TABLE PHONE(
    id VARCHAR(25),
    number VARCHAR(25),
    FOREIGN KEY (id) REFERENCES PERSON(id),
    PRIMARY KEY (id,number)
);
5 Databases Applications

Resources

- http://spots.augusta.edu/caubert/teaching/general/java/
- If you experience troubles, https://www.ntu.edu.sg/home/ehchua/programming/howto/ErrorMessages.html#JDBCErrors might be a good read.
- (Elmasri and Navathe 2010, 13.3.2) or (Elmasri and Navathe 2015, Chapter 10) is a condensed, but good read.
- Many textbook on Java includes a part on Databases, cf. for instance [Gaddis2014, Chapter 16].

5.1 Overview

Two options to interact with a database:

- Interactive interface (C.L.I.), what we used so far
- Application program / Database application
  1. Embed SQL commands in your program: a pre-compiler scans the code, extract the SQL commands, execute them on the DBMS.
  2. Use a library, or Application Programming Interface for accessing the database from application programs.
  3. Create a new language that extends SQL (PL/SQL1)

We will consider option 2. Every database application follows the same routine:

1. Establish / open the connection
2. Interact (Update, Query, Delete, Insert)
3. Terminate / close the connection

<table>
<thead>
<tr>
<th>Python</th>
<th>Python Database API</th>
<th><a href="https://www.python.org/dev/peps/pep-0249/">https://www.python.org/dev/peps/pep-0249/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>C#</td>
<td>MySQL Connector/Net</td>
<td><a href="https://dev.mysql.com/downloads/connector/net/6.10.html">https://dev.mysql.com/downloads/connector/net/6.10.html</a></td>
</tr>
<tr>
<td>Java</td>
<td>Java Database Connectivity</td>
<td><a href="https://docs.oracle.com/javase/9/docs/api/java/sql/package-summary.html">https://docs.oracle.com/javase/9/docs/api/java/sql/package-summary.html</a></td>
</tr>
</tbody>
</table>

5.2 Java’s Way

Java actually uses

- A **protocol** (the API, a class library), Java DataBase Connectivity (JDBC), common to all DBMS. Essentially, a collection of classes to send SQL statements, retrieve and update the results of a query, handle exceptions, etc.
- A **subprotocol** (the driver, connector), Connector/J for MySQL.

And the routine is a bit more complex:

1. Import library
2. Load driver (can also be done at execution time)
3. Open connection (create Connection and Statement objects)
4. Interact with DB (use Statement object)
5. Close connection

5.3 Flash Intro to Java

5.4 A First Program

5.4.1 The Java code

```java
// javac FirstProg.java
// java -cp .:mysql-connector-java-5.1.44/mysql-connector-java-5.1.44-bin.jar FirstProg
import java.sql.*;

class FirstProg {
    public static void main(String[] args) {
        try {
            Connection conn = DriverManager.getConnection("jdbc:mysql://localhost:3306/HW_EBOOKSHOP", "testuser", "password");
            Statement stmt = conn.createStatement();
            String strSelect = "SELECT title, price, qty FROM books WHERE qty > 40";
            System.out.print("The SQL query is: "+ strSelect + "\n");
            ResultSet rset = stmt.executeQuery(strSelect);

            System.out.println("The records selected are:");
            int rowCount = 0;
            String title;
            double price;
            int qty;

            while (rset.next()) {
                title = rset.getString("title");
                price = rset.getDouble("price");
                qty = rset.getInt("qty");
                System.out.println(title + ", "+ price + ", "+ qty);
                rowCount++;
            }

            System.out.println("Total number of records = " + rowCount);
            conn.close();
        } catch (SQLException ex) {
            ex.printStackTrace();
        }
    }
}

A couple of comments:

- java.sql.* contains the classes Connection, Statement, ResultSet, ResultSetMetadata.
- In the string "jdbc:mysql://localhost:3306/HW_EBOOKSHOP", jdbc is the protocol, mysql is the subprotocol, localhost is the url of the database, 3306 is the port, and
HW_EBOOKSHOP is the schema (that needs to already exist in this case).

- Note that strSelect doesn't end with ; (it could, but doesn't have to).
- next() returns true if there is something left in the set of result, and move to the next line if it is the case. It is close to the code we would use to read from a file.
- We could use 1, 2, and 3 instead of "title", "price" and "qty" in the while loop:
  - getString, getDouble and getInt also take integers, corresponding to the position of the attribute in the result set.

### 5.4.2 The database

```sql
/* code/sql/HW_EBOOKSHOP.sql */

CREATE DATABASE HW_EBOOKSHOP;

USE HW_eBOOKShop;

CREATE TABLE BOOKS (  
id int PRIMARY KEY,  
title varchar(50),  
author varchar(50),  
price float,  
qty int  );

INSERT INTO BOOKS VALUES (1001, 'Java for dummies', 'Tan Ah Teck', 11.11, 11);
INSERT INTO BOOKS VALUES (1002, 'More Java for dummies', 'Tan Ah Teck', 22.22, 22);
INSERT INTO BOOKS VALUES (1003, 'More Java for more dummies', 'Mohammad Ali', 33.33, 33);
INSERT INTO BOOKS VALUES (1004, 'A Cup of Java', 'Kumar', 44.44, 44);
INSERT INTO BOOKS VALUES (1005, 'A Teaspoon of Java', 'Kevin Jones', 55.55, 55);

SELECT * FROM BOOKS;
```

MariaDB [HW_EBOOKSHOP]> SELECT * FROM books;

<table>
<thead>
<tr>
<th>id</th>
<th>title</th>
<th>author</th>
<th>price</th>
<th>qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td>Java for dummies</td>
<td>Tan Ah Teck</td>
<td>11.11</td>
<td>11</td>
</tr>
<tr>
<td>1002</td>
<td>More Java for dummies</td>
<td>Tan Ah Teck</td>
<td>22.22</td>
<td>22</td>
</tr>
<tr>
<td>1003</td>
<td>More Java for more dummies</td>
<td>Mohammad Ali</td>
<td>33.33</td>
<td>33</td>
</tr>
<tr>
<td>1004</td>
<td>A Cup of Java</td>
<td>Kumar</td>
<td>44.44</td>
<td>44</td>
</tr>
<tr>
<td>1005</td>
<td>A Teaspoon of Java</td>
<td>Kevin Jones</td>
<td>55.55</td>
<td>55</td>
</tr>
</tbody>
</table>

5 rows in set (0.00 sec)
5.4.3 The result

If you store the program in FirstProg.java, compile it, with

```
javac FirstProg.java
```

and then execute it, with

```
java -cp :mysql-connector-java-5.1.44/mysql-connector-java-5.1.44-bin.jar
    FirstProg
```

then you would obtain:

```
The SQL query is: select title, price, qty from books where qty > 40

The records selected are:
A Cup of Java, 44.44, 44
A Teaspoon of Java, 55.55, 55
Total number of records = 2
```

5.4.4 A variation

If you were to replace the body of try in the previous program with

```
String strSelect = "SELECT * FROM books";
ResultSet rset = stmt.executeQuery(strSelect);

System.out.println("The records selected are:");

ResultSetMetaData rsmd = rset.getMetaData();
int columnsNumber = rsmd.getColumnCount();
String columnValue;
while (rset.next()) {
    for (int i = 1; i <= columnsNumber; i++) {
        if (i > 1) System.out.print(" ", ");
        columnValue = rset.getString(i);
        System.out.print(columnValue + " " + rsmd.getColumnName(i));
    }
    System.out.println();
}
conn.close();
```

You would obtain:

```
The records selected are:
1001 id, Java for dummies title, Tan Ah Teck author, 11.11 price, 11 qty
1002 id, More Java for dummies title, Tan Ah Teck author, 22.22 price, 22 qty
1003 id, More Java for more dummies title, Mohammad Ali author, 33.33 price, 33 qty
1004 id, A Cup of Java title, Kumar author, 44.44 price, 44 qty
1005 id, A Teaspoon of Java title, Kevin Jones author, 55.55 price, 55 qty
```
5.5 Mapping Datatypes

<table>
<thead>
<tr>
<th>SQL</th>
<th>JAVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTEGER</td>
<td>int</td>
</tr>
<tr>
<td>CHARACTER(n)</td>
<td>String</td>
</tr>
<tr>
<td>VARCHAR(n)</td>
<td>String</td>
</tr>
<tr>
<td>REAL</td>
<td>float</td>
</tr>
<tr>
<td>DOUBLE</td>
<td>double</td>
</tr>
<tr>
<td>DECIMAL(t,d)</td>
<td>java.math.BigDecimal</td>
</tr>
<tr>
<td>DATE</td>
<td>java.sql.Date</td>
</tr>
<tr>
<td>BOOLEAN</td>
<td>boolean</td>
</tr>
<tr>
<td>BIT(1)</td>
<td>byte</td>
</tr>
</tbody>
</table>

We cannot always have that correspondence: what would correspond to a reference variable? To a private attribute? This series of problems is called “object-relational impedance mismatch”, it can be overcome, but at a cost.

5.6 Differences Between executeQuery, executeUpdate and execute

<table>
<thead>
<tr>
<th>Name</th>
<th>executeQuery</th>
<th>executeUpdate</th>
<th>execute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used for</td>
<td>SELECT</td>
<td>INSERT, UPDATE, DELETE</td>
<td>Any type</td>
</tr>
<tr>
<td>Input Type</td>
<td>string</td>
<td>string</td>
<td>string</td>
</tr>
<tr>
<td>Return Type</td>
<td>ResultSet</td>
<td>int, the number of rows affected by the query</td>
<td>boolean, true if the query returned a ResultSet, false if the query returned an int or nothing</td>
</tr>
</tbody>
</table>

To retrieve the ResultSet obtained by an execute statement, you need to use getResultSet or getUpdateCount. For more details, consult https://docs.oracle.com/javase/7/docs/api/java/sql/Statement.html.

5.7 A Second Program

5.7.1 Passing options

We can pass options when connecting to the database:

```java
Connection conn =
    DriverManager.getConnection("jdbc:mysql://localhost:3306/HW_DBPROG" + 
    "?user=testuser" + 
    
```
createDatabaseIfNotExist is about schema, actually.

5.7.2 Creating a table
Use stmt.executeUpdate and a create statement. You can use the getMetaData() of the DatabaseMetaData to obtain information about the tables.

5.7.3 Inserting values
Use stmt.executeUpdate (multiple insertion possible if allowMultiQueries was set to true). Another way of batch processing statements:
stmt.addBatch(insert3);
stmt.addBatch(insert4);
stmt.executeBatch();

5.7.4 Prepared statements
“A query with a slot”: parsed and stored on the database, but not executed. When the program gives values, it is executed.
Compared to executing SQL statements directly, prepared statements have three main advantages:

- Reduces parsing time (one time VS as many time as values)
- Minimize bandwidth (send only the parameters, and not the whole query)
- Protect against SQL injections

5.7.5 Advanced Statements objects
You can pass options when creating Statement objects to be able to read it both ways, and to be able to update rows.

COPY HOMEWORK HERE

Exercises

Exercise 5.1 What are the technologies that makes it possible for a Java application to communicate with a DBMS?

Exercise 5.2 What JDBC method do you call to get a connection to a database?

Exercise 5.3 Briefly explain what the next() method from the ResultSet class does, and give its return type.

Exercise 5.4 How do you submit a SELECT statement to the DBMS?

Exercise 5.5 Where is a ResultSet object’s cursor initially pointing? How do you move the cursor forward in the result set?
Exercise 5.6 Give three navigation methods provided by ResultSet.

Exercise 5.7

Explain this JDBC URL format:

```java
jdbc:mysql://localhost:3306/HW_NewDB?createDatabaseIfNotExist=true&useSSL=true
```

Exercise 5.8 In what class is the getColumnName method?

Exercise 5.9 What is a prepared statement?

Exercise 5.10 In the code below, there are five errors between line 13 and line 32. They are not subtle Java errors (like misspelling a key word) and do not come from the DBMS (so you should assume that the password is correct, that the database exists, etc.). For each error, highlight it precisely and give a short explanation.

```java
/* code/java/ProgWithErrors.java */

import java.sql.*;

public class ProgWithErrors{
    public static void main(String[] args) {
        try {
            Connection conn =
                DriverManager.getConnection("jdbc:mysql://localhost:3306/
                +"HW_TestDB?user=testuser&password=password")
                Statement stmt = conn.createStatement();
            }
        }
    /* Errors after this point.*/

        String strSelect = "SELECT title FROM DISKS WHERE qty > 40;";
        ResultSet rset = stmt.executeUpdate(strSelect);

        System.out.println("The records selected are: (listed last
                        first):");
        rset.last();

        while(rset.previous()) {
            String title = rset.getDouble("title");
            System.out.println(title + "\n");
        }

        String sss = "SELECT title FROM DISKS WHERE Price <= ?";
        PreparedStatement ps = conn.prepareStatement(sss);
        ResultSet result = ps.executeQuery();

        conn.close();
    /* Errors before this point.*/
```
Exercise 5.11 Write a small program that determine whenever the null value from Java is equal to the NULL value in your DBMS.

Solution to Exercises

Solution 5.1 API + driver

Solution 5.2 DriverManager.getConnection()

Solution 5.3 It checks if there is data to read, and if move the cursor reads it. It returns a Boolean.

Solution 5.4 Using .executeQuery(strSelect)

Solution 5.5 Before the first line. Using the next method.

Solution 5.6 first, last, next, previous, relative, absolute methods.

Solution 5.7 Connect to localhost:3306 and create a new database if needed, and use secure connection.

Solution 5.8 ResultSetMetaData

Solution 5.9 A prepared statement is a feature used to execute the same (or similar) SQL statements repeatedly with high efficiency.

Solution 5.10 The errors are the following:

- 1. 13: The query is incorrect (WHERE should come before FROM),
- TO be written

Solution 5.11 Here it is:

```java
import java.sql.*;

public class NullProg {
    public static void main(String[] args) {
        try {
            Connection conn = DriverManager.getConnection("jdbc:mysql://localhost:3306/HW_DBPROG?user=testuser&password=password&createDatabaseIfNotExist=true");
            Statement stmt = conn.createStatement();
            stmt.executeUpdate("CREATE TABLE Test (" +
            "A CHAR(25), " +
            "B INTEGER, " +
            "C DOUBLE)");

            String strAdd = "INSERT INTO Test VALUES (NULL, NULL, NULL);";
        }
    }
}
```
int number_of_row_changed = stmt.executeUpdate(strAdd);
System.out.print("This last query changed " +
   number_of_row_changed + " row(s).");

ResultSet result = stmt.executeQuery("SELECT * FROM Test");

if (result.next()) {
    System.out.print(result.getString(1) + " " +
    result.getDouble(2) + " " + result.getInt(3));
    if (result.getString(1) == null) {
        System.out.print("\nAnd null for CHAR in SQL is null for String in Java.\n");
    }
}
conn.close();
} catch (SQLException ex) {
    ex.printStackTrace();
}

5.8 Problems

Problem 5.1 (MySQL’s batch mode and HW_EBOOKSHOP.sql) In the archive, navigate to code/sql/ and open HW_EBOOKSHOP.sql. Then, open a terminal (or command-line interpreter), navigate to the folder where you stored that file (using cd), and type

mysql -u testuser -p < HW_EBOOKSHOP.sql

for linux, or (something like)

"C:\Program Files\MySQL\MySQL Server 5.7\bin\mysql.exe" -u testuser -p
   < HW_EBOOKSHOP.sql

for Windows.

You just discovered MySQL’s batch mode, that perform series of instructions from a file. You can easily make sure that the database and the table were indeed created, and the values inserted.

Problem 5.2 (First Database Application) This exercise supposes you successfully completed Pb 5.1. We will compile and execute your first database application, using Java and MySQL.

- I will assume that you have MySQL installed and set-up as indicated in Homeworks 5.21 and 5.22.
- I will assume that you have Java installed. If not, please refer to http://spots.augusta.edu/caubert/teaching/general/java/ for a simple program and the instructions to compile and execute it.
- We need to set up the driver (or connector) to make the java sql API and MySQL communicate. To do so,
  - Go to https://dev.mysql.com/downloads/connector/j/
5.8 Problems

- Click on “Download” in front of “Platform Independent (Architecture Independent), ZIP Archive”
- Look for the (somewhat hidden) "No thanks, just start my download."
- You will download a file named "mysql-connector-java-***.zip", where *** is the version number.
- Upon completion of the download, unzip the file, and locate the "mysql-connector-java-***-bin.jar" file.
- Copy that file in code/java/.

- Open a terminal in that same folder, and compile FirstProg.java, using javac FirstProg.java (or an equivalent command for windows). Normally, nothing will be printed, but a FirstProg.class file will be created.

- Now, execute that program, using ~~/bash\) java -cp \:mysql-connector-java-***-bin.jar FirstProg ~~/

in Linux, or

\(\) java -cp \:mysql-connector-java-***-bin.jar FirstProg

in Windows. The \(-cp\) option lists the places where java should look for the class used in the program: we are explicitly asking java to use the mysql-connector-java-***-bin.jar executable to execute our FirstProg executable. Try to execute FirstProg without that flag, and see what happens.

Problem 5.3 (Advanced Java Programming) Read, execute, break, edit, compile, patch, hack and (most importantly) understand the following program:

```sql
/* code/java/AdvancedProg.sql */

/*

This is a long program, introducing:

I. How to pass options when connecting to the database,
II. How to create a table
III. How to insert values
IV. How to use prepared statements
V. How to read backward and write in ResultSets

If you want to run this program multiple times, you have to either:

1. Comment first statement of II. Creating a table
2. Change the name of the schema, from HW_DBPROG to whatever you want
3. Drop the DVD table: connect to your database, and then enter
   USE HW_DBPROG;
   DROP TABLE DVD;
   Or do it from within your program!

If you use option 1, you will keep inserting tuples in your table:
   cleaning it with
   DELETE FROM DVD;
   can help. You can do it from within the program!
```
import java.sql.*;

public class AdvancedProg {
    public static void main(String[] args) {
        try (
            Connection conn =
                DriverManager.getConnection("jdbc:mysql://localhost:3306/HW_DBPROG"
                    + "?user=testuser"
                    + "&password=password"
                    + "&allowMultiQueries=true"
                    + "&createDatabaseIfNotExist=true"
                    + "&useSSL=true");

            // Read about other options at
            // https://jdbc.postgresql.org/documentation/head/connect.html

            Statement stmt = conn.createStatement();
        ) {

            /*
             * II. Creating a table
             */

            stmt.execute("CREATE TABLE DVD (" +
                "Title CHAR(25) PRIMARY KEY, " +
                "Minutes INTEGER, " +
                "Price DOUBLE)" +
            );

            /* If we were to execute
             * SHOW TABLES
             * directly in the MySQL interpreter, this would display
             * at the screen
             */
            // +--------------------------+
            // | Tables_in_HW_NewDataBase |
But here, to access this information, we will use the connection's metadata.

```
DatabaseMetaData md = conn.getMetaData();
// DatabaseMetaData is a class used to get information about the database: the driver, the user, the versions, etc.
```

```java
ResultSet rs = md.getTables(null, null, "%", null);
/*
* You can read at
* the full specification of this method.
* All you need to know, for now, is that the third parameter is
* String tableNamePattern,
* i.e., what must match the table name as it is stored in
* the database
* Here, by using the wildcard "%", we select all the table names.
* We can then iterate over the ResultSet as usual:
*/
```

```java
while (rs.next()) {
    System.out.println(rs.getString(3));  // In the ResultSet returned by getTables, 3 is the TABLE_NAME.
}
```

```java
/*
* III. Inserting values
*/
```

```java
String sqlStatement = "INSERT INTO DVD VALUES ('Gone With The Wind', 221, 3);";
int rowsAffected = stmt.executeUpdate(sqlStatement);
System.out.println(sqlStatement + " changed " + rowsAffected + " row(s).\n");
/*
* Batch Insertion
*/
```
String insert1 = "INSERT INTO DVD VALUES ('Aa', 129, 0.2)";
String insert2 = "INSERT INTO DVD VALUES ('Bb', 129, 0.2)";
String insert3 = "INSERT INTO DVD VALUES ('Cc', 129, 0.2)";
String insert4 = "INSERT INTO DVD VALUES ('DD', 129, 0.2)";

// Method 1: Using executeUpdate, if the option allowMultiQueries=true was passed in the url given to getConnection and your DBMS supports it.
stmt.executeUpdate(insert1 + ";" + insert2);

// Method 2: Using the addBatch and executeBatch methods
stmt.addBatch(insert3);
stmt.addBatch(insert4);
stmt.executeBatch();

/*
 * IV. Prepared Statements
 */

// Example 1
sqlStatement = "SELECT title FROM DVD WHERE Price <= ?";
// We have a string with an empty slot, represented by "?".
PreparedStatement ps = conn.prepareStatement(sqlStatement); // We create a PreparedStatement object, using that string with an empty slot.
float maxprice = 0.5f;
ps.setFloat(1, maxprice); // This statement says "Fill the first slot with the value of maxprice".
ResultSet result = ps.executeQuery(); // And then we can execute the query, and display the results:
System.out.printf("For \%2f you can get:\n", maxprice);
while(result.next()){
    System.out.printf("\t %s \n", result.getString(1));
}

// Example 2
sqlStatement = "INSERT INTO DVD VALUES (?, ?, ?)"; // Now, our string has 3 empty slots, and it is an INSERT statement.
PreparedStatement preparedStatement = conn.prepareStatement(sqlStatement);
preparedStatement.setString(1, "The Great Dictator");
p reparedStatement.setInt(2, 124);
5.8 Problems

```java
preparedStatement.setFloat(3, 5.4f);

rowsAffected = preparedStatement.executeUpdate(); // You can check "by hand" that this statement was correctly executed.
System.out.println(preparedStatement.toString() + " changed " + rowsAffected + " row(s).\n");

// If we try to mess things up, i.e., provide wrong datatypes:
preparedStatement.setString(1, "The Great Dictator");
preparedStatement.setString(2, "The Great Dictator");
preparedStatement.setString(3, "The Great Dictator");

// Java compiler will be ok, but we'll have an error at execution time when executing the query. You can uncomment the line below to see for yourself.
//rowsAffected = preparedStatement.executeUpdate();

// Of course, we can use prepared statement inside loops.
for (int i = 1; i < 5; i++) {
    preparedStatement.setString(1, "Saw " + i);
    preparedStatement.setInt(2, 100);
    preparedStatement.setFloat(3, .5f);
    preparedStatement.executeUpdate();
}

/*
* V. Reading backward and writing in ResultSets
*/

// To read backward and write in ResultSets, you need to have a statement with certain options:

Statement stmtNew =
    conn.createStatement(ResultSet.TYPE_SCROLL_SENSITIVE,
    ResultSet.CONCUR_UPDATABLE);

/*
* Those options change two things about the ResultSet we obtain using this statement
* 
* The first argument is the scrolling level:
* TYPE_FORWARD_ONLY = default.
* TYPE_SCROLL_INSENSITIVE = can scroll, but updates don't impact result set.
* TYPE_SCROLL_SENSITIVE = can scroll, update impact result set.
* 
* The second argument is the concurrency level:
*/
* CONCUR_READ_ONLY: default.
* CONCUR_UPDATABLE: we can change the database without issuing SQL statement.
*

/* Reading backward */

sqlStatement = "SELECT title FROM DVD WHERE Price < 1;";
result = stmtNew.executeQuery(sqlStatement);

System.out.println("For $1, you can get:");

if (result.last()) { // We can jump to the end of the ResultSet
    System.out.print(result.getString("Title") + " ");
}

System.out.println("and also, (in reverse order)");

while (result.previous()) { // Now we can scroll back!
    System.out.print(result.getString("Title") + " ");
}

/*
* Other methods to navigate in ResultSet:
* first()
* last()
* next()
* previous()
* relative(x): move cursor x times (positive = forward, negative = backward)
* absolute(x): move to the row number x. 1 is the first.
*/

/*
* Changing the values
*/

System.out.println("\n\nLet us apply a 50% discount. Currently, the prices are:\n\n");

sqlStatement = "SELECT title, price FROM DVD;";
result = stmtNew.executeQuery(sqlStatement);

while (result.next()) {
    System.out.printf("%20s \t $%3.2f\n", result.getString("title") , result.getDouble("price"));
}
result.absolute(0); // We need to scroll back!

while (result.next()) {
    double current = result.getDouble("price");
    result.updateDouble("price", (current * 0.5));
    result.updateRow();
}
System.out.println("\n\nAfter update, the prices are:\n");
result.absolute(0); // We need to scroll back!

while (result.next()) {
    double current = result.getDouble("price");
    result.updateDouble("price", (current * 0.5));
    result.updateRow();
}
System.out.println("\n\nAfter update, the prices are:\n");
result.absolute(0); // We need to scroll back!

while (result.next()) {
    System.out.printf("%20s \t \$%3.2f\n",
        result.getString("title") , result.getDouble("price"));
}
conn.close();
} catch (SQLException ex) {
    ex.printStackTrace();
}
}

Solution to Selected Problems

Solution to Pb 5.2  Without the flag, we obtain:

$ java FirstProg
java.sql.SQLException: No suitable driver found for
  jdbc:mysql://localhost:3306/HW_EBOOKSHOP
at java.sql.DriverManager.getConnection(DriverManager.java:689)
at java.sql.DriverManager.getConnection(DriverManager.java:247)
at FirstProg.main(FirstProg.java:9)
6  A Bit About Security

Resources

6.1  Usual Aspects

6.1.1  Threat model

• Who is threatening you?
• What are the risks?
  1. Loss of integrity (improper modification)
  2. Loss of availability
  3. Loss of confidentiality (unauthorized disclosure)
• “You are as strong as your weakest link.”
• Never trust the user or their computer.

6.1.2  Control measures

• Access control (user account, passwords, restrictions)
• Inference control (cannot access information about a particular “case”)
• Flow control (prevent indirect access)
• Encryption (salting + encrypting, can be a legal obligation): password + salt -> hashed.

Insert short intro. to salting, cryptography.

6.2  Specificities Of Databases

6.2.1  Attack

Attacks: buffer overflow, denial of service, weak authentication, privilege escalation, SQL injections.

“Mixing the instructions with the data”: a judge asking “what is your name”, and you answer "Bill, you are now free to go".

Example with ASP, Active Server Pages, a server-side scripting language:

```plaintext
txtUserId = getRequestString("UserId");
txtSQL = "SELECT * FROM Users WHERE UserId = " + txtUserId;
1. 105; DROP TABLE Suppliers; Execute remote command
2. 105 or 1 = 1 Exploit, bypass login screen
3. admin'-- Line comment, privilege escalation
```

Can also be used for DBMS fingerprinting.
6.2 Specificities Of Databases

6.2.2 Protections

1. Backups:
   
   \texttt{mysqldump --all-databases -u testuser -p password -h localhost > dump.sql}

2. Prepared Statements (a.k.a. stored procedures)

3. White list input validation

4. Escaping

5. Be up-to-date, deactivate the options you are not using, read newsfeeds,
7 Presentation of NoSQL

Resources

7.1 A Bit of History

Inspired from (Sadalage and Fowler 2012, Chap. 1)

7.1.1 DB applications and application DB

When you write a DB application, you have two options:

1. One database for many softwares
2. One database for each softwares

Option a. can cause severe impacts on the efficiency of your database: since maintaining the integrity of the database is a requirement, a lot of synchronization is needed. With option b., you develop an “application database”, and you have more freedom of choice: since only a program interact with a database, you can chose whatever data management you want.

But people were attached to SQL and kept using it.

7.1.2 Clusters, clusters...

Increase in everything (traffic, size of data, number of clients, etc.) meant "up or out", and there was two ways to increase the resources:

1. Bigger machines
2. More machines

Option b. was generally less expensive, but came with two drawbacks w.r.t. databases:

1. Cost of licences,
2. Force to perform “unnatural acts”: relational model are really not made to be distributed
7.1.3 A first shift

- Google Big Table\(^1\), 2004 (made public in ... 2015!) (Chang et al. 2006)
- Amazon DynamoDB\(^2\), 2004 (used in Simple Storage Service (S3) in 2007)
- Facebook’s Cassandra is sometimes mentioned, but it came later on, around 2009 (Lakshman and Malik 2009).

Particular, big company, with specific needs, but people interested in solving some of their problems. Now, people started to think that there could be other ways.

One goal was to get rid of “impedance mismatch”: mapping classes or objects to database tables defined by a relational schema is complex and cumbersome.

Some issues:

- No absolute notion of “private” and “public” in RDBMS (relative to needs)
- Data-type differences (no pointer, weird way of defining string, etc.)
- Value in a relational structure have to be simple (no complex datatype, no structure)

“Impedance mismatch” is that annoying need for a translation.

Also, the data is now

- moving
- growing
- too diverse

for traditional relational DBMS.

7.1.4 Gathering forces

Multiple attempts, going in multiple directions. A meetup to discuss them coined the term “NoSQL” in an attempt to have a “twittable” hashtag, and it stayed (even it is as specific as describing a dog with “no-cat”). The original meet-up asked for “open-source, distributed, nonrelational database”. Today, no official definition, but NoSQL often implies the followig:

- No relational model
- Not using SQL. Some still have a query language, and it resembles SQL (to minimize learning cost), for instance Cassandra’s CQL.
- Run well on clusters
- Schemaless: you can add records without having to define a change in the structure first.
- Open source

Most importantly: polyglot persistence, “using different data storage technologies to handle varying data storage needs.”

7.1.5 The future?

A lot of enthusiasm, also because it “frees the data” (and, actually, the metadata, cf. application/ld+json, JavaScript Object Notation for Linked Data, schema.org, etc.). Some of it will last for sure: polyglot persistency, the possibility of being schema-less, being “distributed first”, the possibility of sacrificing consistency for greater good, etc. Doesn’t mean SQL (“OldSQL”) and

\(^1\)https://cloud.google.com/bigtable/
\(^2\)https://aws.amazon.com/dynamodb/
relational database are over: still useful in many scenario, and the powerfull query language is great (writing your own every time is a nightmare...).

Starting ~ 2010, one reaction was to develop “NewSQL”, which would combine aspects of both approaches. MongoDB announced that it would have more and more of the ACID properties! https://www.mongodb.com/blog/post/multi-document-transactions-in-mongodb

Also, a really great use of NoSQL is to adopt it at an early stage of the development, when it isn’t clear what the schemas should be. When the schemas are final, then you can shift to relational DBMS!

### 7.2 Comparison

#### 7.2.1 Overview

« Comparaison n’est pas raison »

- Semi-structured data (no schema)
- High performance
- Availability
- Data Replication (improves availability and performance)
- Scalability (horizontal scalability (add nodes) instead of vertical (add memory))
- Eventual Consistency
- Natively versionning

Vs

- Immediate data consistency
- Powerfull query language (join is missing from SQL, has to be implemented on the application-side)
- Structured data storage (can be too restrictive)

#### 7.2.2 ACID Vs CAP

ACID is the guarantee of validity even in the event of errors, power failures, etc.

- Atomicity → Transactions are all or nothing
- Consistency → Transactions maintains validity
- Isolation → Executing two transactions in parallel or one after the other would have the same result
- Durability → Once a transaction has been committed, it is stored in non-volatile memory.

CAP (a.k.a. Brewer’s theorem): Roughly, “In a distributed system, one has to choose between consistency (every read receives the most recent write or an error) and availability (every request receives a (non-error) response, without guarantee that it contains the most recent write)” (the P. standing for “Partition tolerance”).

### 7.3 Categories of NoSQL Systems
<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document-based</td>
<td>Data is stored as “documents” (JSON, for instance), accessible via their id (other indexes available).</td>
<td>Apache CouchDB(^3) (simple for web applications, and reliable), MongoDB(^4) (easy to operate), Couchbase(^5) (high concurrency, and high availability).</td>
</tr>
<tr>
<td>Key-value stores</td>
<td>Fast access by the key to the value. Value can be a record, an object, a document, or be even more complex.</td>
<td>Redis(^6) (in-memory but persistent on disk database, stores everything in the RAM!)</td>
</tr>
<tr>
<td>Column-based (a.k.a. wide column)</td>
<td>Partition a table by columns into column families, where each column family is stored in its own files.</td>
<td>Cassandra(^7), HBase(^8) (both for huge amount of data)</td>
</tr>
<tr>
<td>Graph-based</td>
<td>Data is represented as graphs, and related nodes can be found by traversing the edges using path expressions.</td>
<td>Neo4j(^9) (excellent for pattern recognition, and data mining).</td>
</tr>
<tr>
<td>Multi-model</td>
<td>Support multiple data models</td>
<td>Apache Ignite(^10), ArangoDB(^11), etc.</td>
</tr>
</tbody>
</table>

### 7.4 MongoDB

#### 7.4.1 Resources

- [https://oss.sonatype.org/content/repositories/releases/org/mongodb/mongo-java-driver/3.7.0-rc0/](https://oss.sonatype.org/content/repositories/releases/org/mongodb/mongo-java-driver/3.7.0-rc0/)
- [https://jsonlint.com/](https://jsonlint.com/)
- [https://mongodb.github.io/mongo-java-driver/3.4/driver/getting-started/quick-start/](https://mongodb.github.io/mongo-java-driver/3.4/driver/getting-started/quick-start/)
- [https://docs.mongodb.com/manual/administration/security-checklist/](https://docs.mongodb.com/manual/administration/security-checklist/)
- [https://docs.mongodb.com/getting-started/shell/](https://docs.mongodb.com/getting-started/shell/)
- [https://university.mongodb.com](https://university.mongodb.com)
- [https://en.wikipedia.org/wiki/MongoDB](https://en.wikipedia.org/wiki/MongoDB)
- [https://www.w3schools.com/xml/schema_example.asp](https://www.w3schools.com/xml/schema_example.asp)

\(^3\)[https://couchdb.apache.org/](https://couchdb.apache.org/)

\(^4\)[https://www.mongodb.com/](https://www.mongodb.com/)

\(^5\)[https://www.couchbase.com/](https://www.couchbase.com/)

\(^6\)[https://redis.io/](https://redis.io/)

\(^7\)[https://cassandra.apache.org/](https://cassandra.apache.org/)

\(^8\)[https://hbase.apache.org/](https://hbase.apache.org/)

\(^9\)[https://neo4j.com/](https://neo4j.com/)

\(^10\)[https://ignite.apache.org/](https://ignite.apache.org/)

\(^11\)[https://www.arangodb.com/](https://www.arangodb.com/)
7.4 MongoDB

- https://www.w3schools.com/nodejs/nodejs_mongodb_join.asp
- http://api.mongodb.com/
- NoSQL for mere mortal, ch. 6
- NoSQL distilled, ch. 9.

7.4.2 Introduction

MongoDB is

- free (business model: training, support, DB as service, they actually developed MongoDB because they wanted a good solution for a cloud solution!),
- open-source
- cross-platform
- document-oriented (JSON-like documents with schemas).

And there are drivers\(^\text{12}\) for C, C++, C#, Hadoop Connector, Haskell, Java, node.js, PHP, Perl, Python, Ruby, Scala (Casbah).

The mongo shell is an interactive JavaScript interface to MongoDB.

7.4.3 Document-oriented database

Document-oriented database (document store) contains semi-structured data, it is a subclass of the key-value store:

- Relational databases (RDB) pre-define the data structure in the database (fields + data type).
- Key-value (KV) treats the data as a single opaque collection, which may have any number (incl. 0) fields for every record.
- Document-oriented (DO) system relies on internal structure in the data to extract metadata.

RDB is excellent for optimization, but waste space (placeholders for optional values). KV doesn’t allow any optimization, but flexibility and more closely follows modern programming concepts. DO has the flexibility of KV, and allow for some optimization.

One important difference: in RDB, data is stored in separate tables, and a single object (entity) may be spread across several tables. In DO, one object = one instance, and every stored object can be different from every other.

Pro:

- Mapping objects to a DB simpler
- Change “in place”
- Increase speed of deployment

Document:

- Implementations differs on the details of the definition, but always the central notion. MongoDB has its own implementation, but there are ~ 45 others. MongoDB is the most popular one (next: Amazon DynamoDB, Couchbase, CouchDB)
- Documents encapsulate and encode data (Self-Describing Data)
- Do not need to adhere a standard schema.
- One program can have many different types of objects, and those objects often have many optional fields

\(^\text{12}\)https://docs.mongodb.com/ecosystem/drivers/
7.4 MongoDB

- Formats: XML, YAML, JSON, PDF, etc.

MongoDB uses JSON to BSON (portmanteau of the words “binary” and “JSON”), and actually extend JSON. Think of BSON as a binary representation of JSON (JavaScript Object Notation) documents.

An example of XML (Extensible Markup Language) document (you can actually convert from XML to JSON):

```xml
<shiporder orderid="889923">
  <orderperson>John Smith</orderperson>
  <shipto>
    <name>Ola Nordmann</name>
    <address>Langgt 23</address>
    <city>4000 Stavanger</city>
    <country>Norway</country>
  </shipto>
  <item>
    <title>Empire Burlesque</title>
    <note>Special Edition</note>
    <quantity>1</quantity>
    <price>10.90</price>
  </item>
  <item>
    <title>Hide your heart</title>
    <quantity>1</quantity>
    <price>9.90</price>
  </item>
</shiporder>
```

- Invalid document exists!
- Human and computer-readable
- No predefined tags
- Extensible

### 7.4.4 General organization of MongoDB databases

<table>
<thead>
<tr>
<th>RDBMS</th>
<th>MongoDB</th>
</tr>
</thead>
<tbody>
<tr>
<td>database instance</td>
<td>MongoDB instance</td>
</tr>
<tr>
<td>schema</td>
<td>database</td>
</tr>
<tr>
<td>table</td>
<td>collection</td>
</tr>
<tr>
<td>row</td>
<td>document</td>
</tr>
</tbody>
</table>

Each MongoDB instance has multiple databases, each database can have multiple collections.

Two documents (delimited by [...], used to delimit an array of document):

```json
[
  {
    "firstname": "Martin",
    "likes": [ "Biking",
```
"Photography" ],
"lastcity": "Boston",
"lastVisited":
}
,
{
"firstname": "Pramod",
"citiesvisited": [ "Chicago", "London", "Pune", "Bangalore" ],
"addresses": [ 
  { "state": "AK",
    "city": "DILLINGHAM"
  },
  { "state": "MH",
    "city": "PUNE"}
],
"lastcity": "Chicago"
]}

Use the same example all along!

- addresses is a document embedded in a document!
- Some attributes are common, some are not: that’s fine, every document can have its own schema.

A collection should be on “related” entities (do not store server logs, store customers and list of employee in the same collection!), and not too abstract ones (no “Server stuff”). Also, if you store document that are too different, your performances will take a big hit. Bottom line: think about your usage, and the kind of queries you will perform.

"Schema-less" does not mean “organization-less”!

### 7.4.5 First elements of syntax

```javascript
db.<name of the collection>.<command> (db is not the name of the database, it is just the prefix).
db.book.insert({"title": "Mother Night", "author": "Kurt Blabal"})

MongoDB will add a unique identifier (_id) if you do not provide one. You can think of that as a primary key.

db.books.remove({"title":"Mother Night"})
db.books.update({"title":"Mother Night"}, {$set {"quantity" : 10}})

Other function, such as $inc, to increment.

db.books.find() is like SELECT * FROM Books;.
db.books.find({"title":"Mother Night"})
db.books.find({"title":"Mother Night"}, {"author":1, "quantity":1})
db.books.find({"title":"Mother Night"}, {"author":0, "quantity":0}): everything but the author and the quantity.
```
db.books.find({"quantity":{"$gte": 10, "$lt": 50}}): greater than equal to 10, less than 50.

Possibility to mimic some features (unique attributes), but no referential key integrity, for instance.

Most insert / update / delete will return success as soon as one node received your command, but you may tweak them so that success is returned only once the operation has been performed on the majority of the nodes.

### 7.4.6 Set up

Install `mongodb` (non-official version, not maintained by MongoDB) and download [https://mongodb.github.io/mongo-java-driver/](https://mongodb.github.io/mongo-java-driver/).

Command-line: `mongo`

Not a lot of features, need to write a lot on the program side. But there are tons of API (“package manager” approach), cf. for instance an API over mongo-java-driver: [http://jongo.org/](http://jongo.org/) (support some form of prepared statement)

The documentation is nicely written and well-organized: we’ll follow parts of it, please refer to it if needed.

### 7.4.7 Example


Compile and execute with

```bash
javac -cp .:mongo-java-driver-3.7.0-rc0.jar MongoTest2.java
java -cp .:mongo-java-driver-3.7.0-rc0.jar MongoTest2
```

After various import statement, and the usual header:

```java
MongoClientURI connectionString = new
MongoClientURI("mongodb://localhost:27017");
MongoClient mongoClient = new MongoClient(connectionString);
```

Or, more compact:

```java
MongoClient mongoClient = new MongoClient();
```

Get a collection:

```java
MongoDatabase database = mongoClient.getDatabase("mydb");
MongoCollection<Document> collection = database.getCollection("test");
```

Assume we want to create the following document:

```java
{
    "name" : "MongoDB",
    "type" : "database",
    "count" : 1,
    "versions": [ "v3.2", "v3.0", "v2.6" ]
}```
7.5 Principles

• “Schemaless means more responsibility”
• Some denormalization, sometimes: duplicate the information, to have it all in one place. Example: table for phone number, for employee, for emergency contact. You can duplicate that information, no big deal. Less join (resources expensive), but need more storage, more functions, to substitute.
• NoSQL injection: your application should accept only strings from your users (never allow objects by design) and sanitize the inputs before using them (mongo-sanitize is a good module for this).

Exercises

Exercise 7.1 What is polyglot persistence? Is it useful?
Exercise 7.2 What does it mean to be “schemaless”? What does it imply?
Exercise 7.3 What is denormalization? When could that be useful?

Exercise 7.4 What is the (object-relational) impedance mismatch? Is it an issue that cannot be overcome?

Solution to Exercises

Solution 7.1 It is the task of picking the right DBMS for the task, and to involve multiple DBMS in a single application. Yes, it is useful. Per wikipedia, "Polyglot persistence is the concept of using different data storage technologies to handle different data storage needs within a given software application."

Solution 7.2 That a table can contain documents, or tuples, with different attributes. It implies more responsibilities.

Solution 7.3 To duplicate data about other entities in some entities. It is useful when joining is expensive.

Solution 7.4 Data-base and object-oriented principles are different and it requires work to make them work together. This correspondance, or matching, can be implemented in the application, or lead to the design of new DBMS.

13https://en.wikipedia.org/wiki/Polyglot_persistence
References


Sullivan, Dan. 2015. NoSQL for Mere Mortals. Addison-Wesley Professional.