

COMP23111: Fundamentals of Databases Practical Sessions Manual

2014-2015 edition

Alvaro A A Fernandes, Sandra Sampaio, Klitos Christodoulou
School of Computer Science
University of Manchester

Introduction	6
<i>Aims</i>	<i>6</i>
<i>General Approach</i>	<i>6</i>
<i>On Plagiarism and Working Together</i>	<i>6</i>
General Information	7
<i>Communication and Announcements</i>	<i>7</i>
<i>Oracle Usernames and Passwords</i>	<i>8</i>
<i>Attendance</i>	<i>8</i>
<i>Timetable</i>	<i>8</i>
Lab Sessions/Example Clinics	8
<i>Specifications</i>	<i>8</i>
<i>Marking and Feedback</i>	<i>9</i>
<i>Lab Exercise Deadlines</i>	<i>9</i>
<i>Extensions to Lab Exercise Deadlines</i>	<i>9</i>
<i>Submission</i>	<i>10</i>
<i>The Orinoco Case Study Used in Lab Exercises</i>	<i>11</i>
<i>Interpreting Examples, Requirements and Specifications</i>	<i>11</i>
Examples Clinic 1	13
<i>Understanding the Relational Paradigm</i>	<i>13</i>
<i>Specification</i>	<i>13</i>
Lab Exercise 01	15
<i>First Steps with SQL *Plus and Dia for (E)ER Modeling</i>	<i>15</i>
<i>Specification</i>	<i>15</i>
<i>Overview</i>	<i>15</i>
<i>Introduction</i>	<i>15</i>
<i>Requirements</i>	<i>16</i>
<i>Instructions</i>	<i>16</i>
<i>Submission Contents</i>	<i>16</i>

Part 1: SQL*Plus TUTORIAL	16
Part 2: Getting Started with Dia	24
Examples Clinic 2	26
<i>Entity-Relationship Modelling</i>	26
<i>Specification</i>	26
Lab Exercise 02	28
<i>(E)ER Modelling</i>	28
<i>Specification</i>	28
<i>Overview</i>	28
<i>Introduction</i>	28
<i>Requirements</i>	28
<i>Instructions</i>	28
<i>Submission Contents</i>	29
<i>Part 1: Understanding the ER schema for the partial Orinoco database</i>	29
<i>Part 2: Extending the ER schema for the Orinoco database</i>	30
Examples Clinic 3	31
<i>Mapping Entity-Relationship Diagrams into Relational Schemas</i>	31
<i>Specification</i>	31
Lab Exercise 03	32
<i>EER-to-R Mapping, Functional Dependencies, SQL DDL/DML</i>	32
<i>Specification</i>	32
<i>Overview</i>	32
<i>Introduction</i>	32
<i>Requirements</i>	32
<i>Instructions</i>	32
<i>Submission Contents</i>	33
<i>Part 1: Applying the EER-to-relational mapping to the extended Orinoco database</i>	33

<i>Part 2: Updating the schema/populating new relations in the Orinoco database</i>	33
Examples Clinic 4	36
<i>Normalising Relational Schemas</i>	36
<i>Specification</i>	36
Lab Exercise 04	37
<i>Making Use of a Relational Database</i>	37
<i>Specification</i>	37
<i>Overview</i>	37
<i>Introduction</i>	37
<i>Requirements</i>	37
<i>Instructions</i>	37
<i>Submission Contents</i>	37
<i>Part 1: Writing simple queries as RA expressions and translating them into SQL</i>	38
<i>Part 2: Writing and executing complex SQL queries</i>	38
<i>Part 3: Using SQL queries to model data requirement scenarios</i>	39
Examples Clinic 5	40
<i>Advanced SQL and Transactions</i>	40
<i>Specification</i>	40
Lab Exercise 05	42
<i>Advanced SQL: Triggers and Procedures</i>	42
<i>Specification</i>	42
<i>Overview</i>	42
<i>Introduction</i>	42
<i>Requirements</i>	42
<i>Instructions</i>	42
<i>Submission Contents</i>	43
<i>Part 1: Writing triggers</i>	43

Part 2: Writing procedures

43

Introduction

Aims

This course unit aims to introduce the fundamentals of databases from one particular viewpoint, viz., that of business information systems, where modern database management systems play a pivotal role.

The lectures cover key topics in the area of database design and implementation while the practical sessions, which are the focus of this manual, provide substantial illustration of the topics covered, allowing you to apply and practice some of the notions and concepts introduced in the lectures.

Note that, in this course unit, we refer to practical sessions that are assisted, but not assessed, as *example clinics* because, although we recommend attendance (and record it), they are not strictly compulsory, as no marks are available in these sessions. However, if we slip into calling them *example classes*, do not feel confused, as we do not mean yet something else.

General Approach

The lab exercises involve the design and construction of a fairly complete database from a hypothetical case study. The sequence of lab exercises will be seeded by a partially completed database implementation that is provided to you.

The lab exercises are conducted in five stages. The first stage is mainly an introduction to the tools you need to use, and therefore contains little in the way of problem-solving. The following four, in contrast, contain more problem-solving tasks.

It is very important that you bear in mind that the case study is *incremental* in the sense that later lab exercises build on **your own answers** to earlier ones. However, and very importantly, model answers are not handed out after each exercise (i.e., there is no reset).

This means that you must yourself be aware of any deviation from the right track (broadly sketched out for you in the feedback you get from each exercise) and, if needed, fix your solution thus far. In other words, from the start, think of the sequence of exercises as, in fact, a single task comprising several incremental stages. This mirrors real-world database design and development, which is a multistage process that unfolds over a period of months, rather than days or weeks.

When the coursework has been marked, we aim to give you such feedback as will help you fix any problems in your submission, so that the subsequent exercise is done from a correct intermediate position. If you want to discuss the feedback we give you, please use the various mechanisms available, viz., the forum (for general questions) or a chat with us at a lab session, or else, for more specific and personalised feedback, use a direct email to the lecturers or the regular weekly one-hour open-office slot (see details of that in [the Moodle webpage for the course unit](#)).

On Plagiarism and Working Together

You must always work individually. You must only consult and discuss things with your friends once you have gained a thorough understanding of what constitutes academic malpractice. You are not allowed to copy solutions from anyone else, or to pass solutions to anyone else, in any form (conversation, paper, electronic media, etc.), unless otherwise authorised explicitly.

If you have not done so yet, make absolutely sure that you understand and abide by the guidance on plagiarism and cheating:

<http://studentnet.cs.manchester.ac.uk/assessment/plagiarism.php>

You should remind yourself of, and be wise to follow, John Latham's thoughts on the issue of students working together. Here is a reminder (very lightly adapted) for your convenience:

Many of you will be more than tempted to work together with your friends on the solutions to the problems. We have no objection to this in principle, but be aware of the fine line between working together and copying. We will not tolerate copying. If you do not actually do, on your own, with full understanding, every part you are supposed to do of each exercise that you get marked, then you have copied. Remember the requirement that you fully understand your work, and our reserved right to viva you on it without notice!

When working together in an informal group of friends, pay special attention to the relative abilities of people in that group. The real point of laboratory work is not to get the answer, but to learn by doing it yourself, even if you actually get the wrong answer! If you find yourself always telling your friends the answers, then you are not much of a friend to them! You are holding them back and patronising them, but more to the point, you are undermining their learning. Equally, if you find yourself with a friend who keeps telling you the answers, don't be grateful! There is a good chance that he, or she, is simply trying to impress you. Don't be impressed!

Everybody who works in an informal group should actually do the work themselves, individually. Such working together should be restricted to discussing ideas and getting the work off the ground. Anybody who cannot actually do the work, should get help from one of the TAs¹ or supervisors. These people are experienced at helping you in such a way that you can do the work yourself, rather than just giving you the answer.

In determining the line between help and plagiarism, a good rule of thumb to stick to is this: if you get help from another student, or give help to another student, make sure that the passage of information between you is at a much higher level of abstraction than the final answer, no matter what the medium is (soft copy, typed, written, spoken, etc.). So, for example, for exercises which involve you writing some program code (and which is not supposed to be team work!), never show your code, or your detailed pseudocode, to another student: that amount of "help" is cheating. Thus, for example, if you and a few friends develop a solution to a laboratory exercise together, you are, by definition, cheating.

On the other hand, you can help each other as much as you like about things that you have not been asked to create, such as explaining the exercise question to each other, or explaining material that has been covered in lectures, e.g., what is a foreign key in the context of the relational model, etc..

Read and heed the above. To reiterate once more: we do not tolerate copying.

General Information

Communication and Announcements

Communication between you and the lecturers and TAs is primarily via Moodle, so it is important that you check [the Moodle webpage for the course unit](#) daily, and certainly at least once or twice a week. Also, check your School and University email inboxes and check the course unit forum in Moodle, which is accessible from the course unit webpage. Of course, for matters specifically related to you alone, emailing us is fine.

It is your responsibility to keep an eye for our attempts to communicate with you. Check for any such attempt often, read them carefully and respond promptly if a response is required.

¹ TA stands for Teaching Assistant, who used to be known as Demonstrator in the past.

Oracle Usernames and Passwords

You will be using the Oracle DBMS (and, in particular, its command-line interface, Oracle SQL*Plus) in lab exercises (as well as for self-learning). For that, you will need an Oracle username and password.

You are assigned these and they are sent to you by email. If you have not received yours by the start of Week 3, email the TA supervisor and both lecturers to let them know ([their emails are available in the course unit webpage](#)).

Attendance

Attendance is mandatory for lab sessions and very highly recommended for examples clinics. If you miss one of those, it is your responsibility to ensure (as soon as possible, and typically through a fellow student) that you are aware of any announcements made during the session and that you obtain a copy of any handouts and similar material provided during the session.

Attendance is taken for all lab session and examples clinics. The session/clinic supervisor will attempt to see everyone and take their attendance, but it is your responsibility to make sure you have had your attendance taken.

Besides the basic fact that you need to practice in order to learn, your attendance record may become important if you need an extension (see below) or to plead for special treatment in case something goes wrong for you. If you have a good attendance record, your case is likely to be far stronger than if you do not.

Timetable

The [timetable](#) for the whole of the course unit is available in the School's StudentNet.

The course unit webpage provides further information on a week-by-week basis as the course unit is being delivered.

Lab Sessions/Example Clinics

All lab exercises last for one lab session only (see below for deadlines and for the policy on extensions). Lab exercises are to be undertaken individually but examples-clinic exercises can be worked on in small groups.

As usual, lab sessions take place in A-weeks (except for Week 1) and examples clinics take place in B-weeks. Note that the week after Reading Week is Week 7, a B-week.

Examples clinics prepare you for the subsequent lab session/exercise. However, you can use examples clinics to ask for ongoing help and face-to-face, one-to-one feedback on any aspect of the syllabus, so do take advantage of this if you find yourself in need of more clarification or more specific help.

Sample answers are provided for the examples in examples clinics, typically on the week after they were held. However, for reasons explained above, sample answers are not provided for lab exercises, only individual feedback. Since the lab exercises are incremental, this has important consequences that you need to manage yourself.

Specifications

The specifications and detailed instructions for lab exercises and examples clinics are available below.

In both lab and examples-clinic exercises, some of the questions are, as you would expect, harder than others. Try your best and if you cannot quite answer them, do not be disappointed: they are there for us to understand how much you can figure out for yourself using self- as well as group-learning. When it happens that you cannot answer, ask for help and the lecturer/TAs will do their best to help you work it out.

Marking and Feedback

The distribution of marks is spelled out in the lab specifications below.

Note that some marks are set aside for good report presentation. Broadly speaking, this means: correct files, correct file names, correct file format (i.e., PDF), proper formatting (including the use of indentation and fixed-width fonts when presenting formal text), and clear grammatical English.

It is important, therefore, that you develop the discipline to produce a report for assessment that conforms in terms of layout and formatting, clearly presented in terms of language, and clearly explained in terms of the substance of your answers.

An important aspect of good report presentation is legibility of the results printed by Oracle SQL*Plus. You are shown how to control the width of columns, for example. In general, you are required to use the controls at your disposal in Oracle SQL*Plus to make the submission abide by the good presentation criteria that have just been discussed.

All coursework is marked offline but the lecturers and TAs will be present in the lab to help you learn and to give you feedback as you require, including on the marking.

We are committed to returning marks and feedback as promptly as possible. If you feel that you need more feedback, or more information on the feedback you have received, approach us in a lab session, where we will all be ready to talk to you about it.

All lab exercises are worth up to 20 marks. Marks may be scaled automatically school policy dictates so.

Lab Exercise Deadlines

Every lab exercise is a one-session exercise. Therefore, the time you have for completing each lab exercise is the same for all five of them: a single lab session. This means that, **the Arcade-enforced deadline for every lab exercise is the same: the end of the corresponding lab session.**

It follows from this that we assume you will arrive at the lab session with full understanding of the lab exercise specification and, as a rule, with some, even most, of the work already done. The lab session is therefore mostly to help you with face-to-face feedback on difficulties you have encountered when trying to do the exercise in advance of attending the session.

If you fail to submit your work by the end of the session, technically speaking you need an extension.

Extensions to Lab Exercise Deadlines

For all lab exercises in this course unit, if you have been in the lab working on the exercise from start to end and if you have worked on it in a sustained and focussed manner, you can ask for an extension.

In terms of the general School policy, an extension to a lab exercise deadline covers the period from the end of the corresponding lab session either to the starting-time-of-the-lab-session one week after the day of the lab session or to the beginning of the next scheduled lab session. Whichever of these is the shorter period is then the period covered by the extension. Arcade automatically enforces this rule.

For example, when lab sessions happen every fortnight (as is the case for COMP23111), for a Thursday 11am-13am lab session, the corresponding lab exercise's extended deadline is the following Thursday (i.e. 7 calendar days later) at 11am.

Extensions are dependent on attendance and on your working on the exercise throughout the lab session. So, the lab supervisor will take attendance **10 mins after the start of the session**, if you are not present then, you will not be granted an extension unless you can prove that there were special circumstances causing you to be late or that have made it impossible for you to attend at all.

The lab supervisor will accept requests for extensions starting **20 mins before the end of the session**. If you need to leave early, you must have a compelling reason to give to the lab supervisor.

If your attendance is taken *and* you have spent all time working on the exercise *and* you ask for an extension, the lab supervisor will grant it.

Beware: if you spend any time in any other activity without having already submitted your work, you will get a warning. If you persevere in this practice, you will not be granted an extension.

If you miss the extended deadline, your submission incurs the usual late penalty.

No further extensions are available, so refrain from asking for one unless you have special circumstances to report. If you do, then both follow the proper procedure described in the School's StudentNet *and* alert the TA supervisor and the lecturers in person or by email at the earliest possible opportunity: don't leave any special circumstances unreported for longer than strictly necessary.

Submission

The submission is through Moodle, so look for the appropriate submission link in the Moodle course unit webpage. Note, though, that it is Arcade that will determine whether you have met the deadline or not. In the latter case, any penalty is applied by Arcade.

It is absolutely crucial that you study the `ReadMe.txt` file provided in:

`/opt/info/courses/COMP23111/lab_submission_N.zip`

We'll lightly summarize the main points here, but you **MUST** study it otherwise you'll have problems submitting your coursework and getting it marked fairly and correctly.

For all exercises, you are required to produce a report as well as generate and then submit some SQL scripts along with the spooled out result of running those scripts in Oracle SQL*Plus.

In order to do that, you will likely work on a text editor, copying-and-pasting SQL statements and SQL*Plus commands, then debugging them until you get them right. Once you have worked through the lab specification, debugged everything and are convinced that you have composed a script that produces the right output for submission, you **must** use the `SP00L` command in Oracle SQL*Plus (see Lab Exercise 1) to redirect the formatted output of the script onto a file named `LabExNOut.lst`, where `N` is the lab exercise number. Then, that `.lst` file will go into the LaTeX document to make it part of your report. Read the `ReadMe.txt` file we mentioned above.

Avoid re-typing or adding anything to the `.lst` file created by running the script since you could insert errors in it. Be very careful to avoid disturbing indentation and alignment, because it leads to unreadability, which, in turn, incurs a loss of marks.

Keep all the files (i.e., diagrams, scripts, listings, etc.) that went into your report safe and well-organised in the appropriate folders as described in the `ReadMe.txt` file we mentioned above.

Unless you have a good reason for it, do not do anything that alters the timestamps, as this would invalidate their status as evidence if we were to ask you to show them to us when we audit your submission.

You are also required to submit the `.dia` file of diagrams we ask you to generate. For inclusion in your report, a Dia diagram should be saved in Encapsulated PostScript (EPS) format. Read the `ReadMe.txt` file we mentioned above.

You must submit your report through Moodle, and only through Moodle, using the appropriate links that are available for this purpose in the course unit webpage.

For each lab exercise, when you submit your coursework for marking, Moodle will ask you to submit **two** (and **only two**) files, as follows:

1. A report, which is what we will mark. This must be written in LaTeX, using the template we provide (see above) to produce a PDF document. You *don't submit the .tex file* separately, rather **you must submit the PDF file** generated as explained in the `ReadMe.txt` file we mentioned above.
2. A .zip file that is a compressed archive of the folder where you will organize the outcome of your work on a given lab exercise and that we will need for reference purposes in marking the report and for auditing malpractice, if needed.

The template will enforce some rules, as follows. The first page of the report must be a title page which should clearly state your name, your student id, and the lab exercise number. You must have a header consisting of your student ID and the lab exercise number and a footer that causes all pages to be numbered in the format *x of y* (e.g., 2 of 6).

The report file should be named as follows:

COMP23111_14_Lab_Exercise_<n>_<firstName>_<lastName>.pdf

where <n> is the sequence number of the lab exercise, and <firstName> and <lastName> are your first name and your surname, respectively. Make sure that your first name and surname are consistent with how Campus Solutions has recorded your personal data.

The *Orinoco* Case Study Used in Lab Exercises

The example clinics will use several, different short scenarios from different application domains but the lab exercises will centre around the following single, hypothetical case study, the preamble of which is as follows:

Orinoco is a medium-sized record producer that operates on niche markets and offers music on CDs, tape, and vinyl. The firm has recently bought a number of smaller recording companies and is planning to expand.

There is a great deal of information that remains available only in the brochures that *Orinoco* produces, along with an accompanying price guide. Some information is stored on electronic files on an ad hoc basis. This situation has now become very cumbersome as such files lie in different computers, leading to inconsistency and mutual incompatibility. As a result, from the point of view of the software applications that the company aims to develop to underpin its expansion plans, the *Orinoco* data assets seem incomplete and are difficult to exploit to the full (particularly due to the great heterogeneity injected by the lack of proper assimilation of the data assets that came with the recent takeovers). For example, it is particularly difficult to find music that matches a customer's specific requests unless brochures are browsed carefully, which is time-consuming and error-prone.

The management at *Orinoco* have decided to computerise the information in their brochures, to reconsider the data stored in files, to include purchase information, and to sell directly over the web. There is a need, therefore, to make it easier for software applications to:

1. Gather specific information from customers and answer questions about products
2. Check the availability of specific products
3. Give costs of products
4. Keep information about artists, albums, and tracks
5. Keep information about the back catalogues of the companies recently taken over
6. Keep contract information about artist signings etc.

You may assume that the system is stand-alone and does not have to integrate with any other. For the purposes of this course unit, other simplifying assumptions have been made (e.g., that price information is fixed and is not flexible, etc.).

Interpreting Examples, Requirements and Specifications

You **must** solve the problems as specified in the instructions you are given, not as you imagine that they would be in your personal view of what the real world is like.

Ignoring the stated requirements because, as a designer, you feel you know best what is correct or what is needed is an outright mistake.

In a professional context, it would be a case of misconduct to ignore or override customer requirements at will. In the context of this course unit, you will lose marks and end up getting yourself in difficulties (remember that each exercise builds on the preceding ones without any reset).

Of course, if the requirements are incomplete and ambiguous, which would lead, in the real world, to further rounds of elicitation requirement interactively with the customers. In this course, you should instead consult the TA supervisor or the lecturers.

In other words, do not go beyond or distort what is stated as the user requirements. If you feel they are unclear, check with us.

Examples Clinic 1

Understanding the Relational Paradigm Specification

The goal of this examples clinic is to help you improve your understanding of the relational model, the relational algebra and basic SQL. It aims to prepare you for Lab Exercise 1, in which you get the first opportunity to use the Oracle DBMS.

DURATION: 50 minutes examples clinic time.

DESCRIPTION: The examples clinic consists of five tasks, which are questions on the relational model, relational algebra and SQL. Work through the questions, show them to the lecturers and TAs in order to obtain feedback.

TASKS

T1. Consider the following relational schema. The context is banking, where banks, in the usual way, have branches with accounts assigned to customers. Customers make deposits on these accounts and take loans when they need. Note that this schema may seem odd to you when you start pondering it more deeply. For example, if you were in charge of this bank, you might go about the business of ensuring that every relation has a primary key in a rather different way than they seem to have done. However, remember, this is the bank's stated view. So, assume it is intended. Now, propose appropriate primary keys by underlining the participating attributes:

```
branch(branch name, branch city, assets)
customer (customer name, customer street, customer city)
loan (loan number, branch name, amount)
borrower (customer name, loan number)
account (account number, branch name, balance)
depositor (customer name, account number)
```

T2. Given your choice of primary keys in **Q1**, identify appropriate foreign keys and state what primary key each foreign key references.

T3. Consider the UNIVERSITY relational schema in diagrammatic form shown in Figure EC-01.

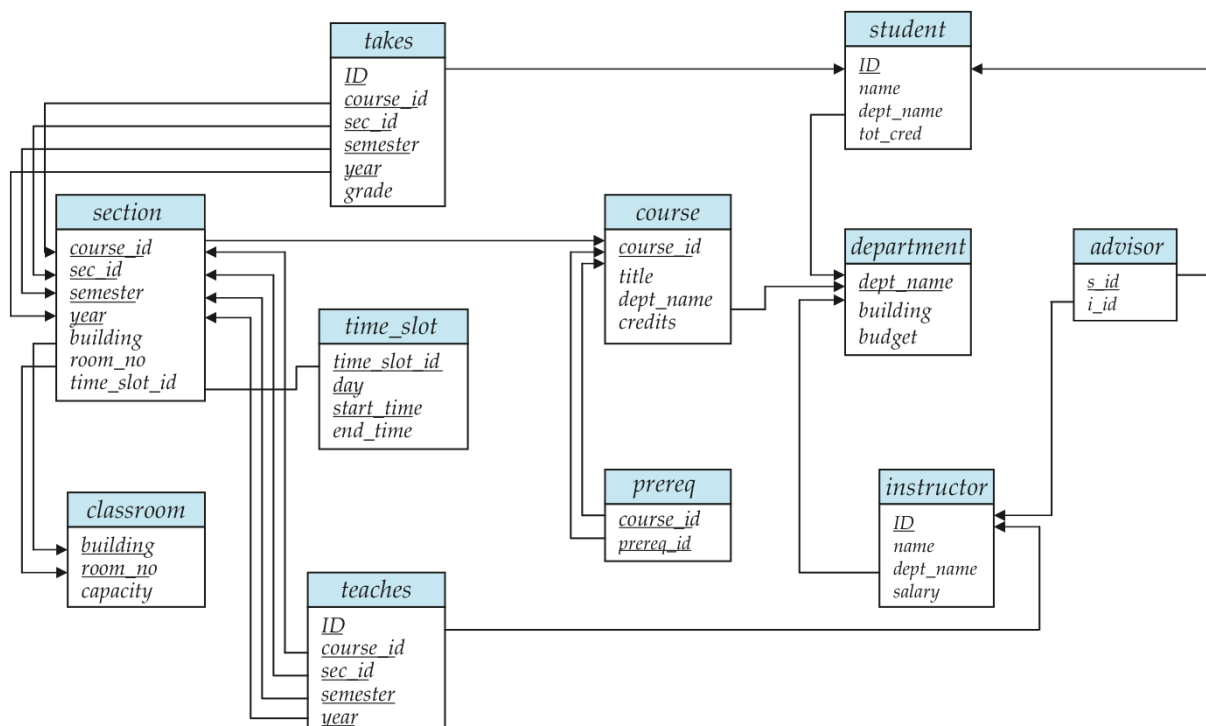


Figure EC-01: UNIVERSITY Relational Schema in Diagrammatic Form

This diagrammatic form denotes a relation with a rectangle, the relation name (e.g., *takes*) appears shaded/light-blue, the attributes are below the relation name (e.g., *ID*, *course_id*, *sec_id*, *semester*, *year*, *grade*), with the attributes that participate in the primary key underlined. An arrow from an attribute *a'* in relation *S* to an attribute *a* in relation *R* denotes that the latter occurs as a foreign key in the former. For example, there is an arrow from *takes.ID* to *student.ID* so you can read the arrow as '*takes.ID* references *student.ID*'.

Now, consider the *advisor* relation, with *s_id* as its primary key. Suppose a student can have more than one advisor. Explain why would *s_id* not be an appropriate primary key of the *advisor* relation any more? What would an appropriate primary key for the relation be in that case?

T4. In the context of the UNIVERSITY schema given above for **Q3**, write the following queries in SQL:

- I. Find the names of all students who have taken at least one Computer Science course; make sure there are no duplicate names in the result.
- II. Find the IDs and names of all students who have not taken any course offering before Spring 2009.
- III. For each department, find the maximum salary of instructors in that department. You may assume that every department has at least one instructor.
- IV. Find the lowest, across all departments, of the per-department maximum salary computed by the preceding query.

(Note that you may need more SQL constructs than were discussed in the lectures. This is a good reminder that you need to read the mandatory readings.)

T5. (Note that we may not have yet covered relational algebra in the lectures at this point. If so, you may prefer to leave this final question to answer later, when you have caught up on the mandatory readings.)

Consider the following relational schema:

```
employee (person name, street, city)
works (person name, company name, salary)
company (company name, city)
```

Work out what are the primary and foreign keys, then, for each of the following data requirements, give a relational algebraic expression that captures them:

- I. Find the names of all employees who work for "First Bank Corporation".
- II. Find the names and cities of residence of all employees who work for "First Bank Corporation".
- III. Find the names, street address, and cities of residence of all employees who work for "First Bank Corporation" and earn more than \$10,000.

Acknowledgements

The tasks in this practical session are minor adaptations of material authored and made available to instructors by A. Silberschatz, H. F. Korth, and S. Sudarshan to accompany their textbook *Database System Concepts*, 6th Edition McGraw-Hill, 2006, 978-0-07-352332-3 and by H.P. Garcia-Molina, J.D. Ullman, and J. Widom to accompany their textbook *Database Systems: The Complete Book*, 2nd Edition Pearson Prentice Hall, 2009, 978-0-13-135428-9. Copyright remains with them, whom we thank. Any errors are our responsibility.

Lab Exercise 01

First Steps with SQL*Plus and Dia for (E)ER Modeling Specification

Overview

The goal of this lab exercise is to provide you with an opportunity to interact with the Oracle DBMS using Oracle SQL*Plus, the command-line interface provided by Oracle, and with the Dia program for drawing structured diagrams.

You are given scripts that allow you to create/drop/recreate some tables from the *Orinoco* case-study database and populate them with tuples. Using SQL*Plus, this then allows you to explore how SQL is implemented by Oracle.

You will gain a basic understanding of the data dictionary component of the Oracle DBMS and will take the initial steps in using the data definition and manipulation aspects of SQL, as implemented in Oracle, as well as basic query forms. You will also gain an understanding of how referential integrity is implemented.

Finally, you will be asked to reproduce an (E)ER diagram using Dia, just to ensure that you become familiar with its support for (E)ER modelling.

ATTENDANCE: Compulsory
 MARKS AVAILABLE: 20 marks.
 DEMONSTRATION: You are required to submit two files: a report in PDF format generated from the template we provide and an archived folder of the outcomes of your work (scripts, spooling output, and diagrams).
 TIME AVAILABLE: 2 hour lab session.
 DEADLINE: The end of the session.
 EXTENSIONS: Provided you attended the lab session and worked on the exercise all the time, you are granted an automatic extension. No further extensions are granted without documented special circumstances. This is strictly enforced.

Introduction

This exercise has two parts: a tutorial followed by an exercise.

You are given information on how to run some scripts that will result in your being the owner of the following populated tables from the *Orinoco* database.

Manager	(<u>manager_ID</u> , name)
ManPhone	(<u>manager_ID</u> [fk:Manager.manager_ID], telephone)
Artist	(<u>artistic_name</u> , genre, managedBy [fk:Manager.manager_ID])
Album	(<u>album_ID</u> , title, createdBy [fk:Artist.artistic_name])
MasterTrack	(<u>track_ID</u> , working_title, working_length, recordedBy [fk:Artist.artistic_name], editedBy [fk:SoundEngineer.sound_eng_id])
SoundEngineer	(<u>sound_eng_ID</u> , name)
ContractInfo	(hasContract [fk:Artist.artistic_name], date_from, date_to, duration [derived as (date_to - date_from)])
FinishedTrack	(<u>originatesFrom</u> [fk:MasterTrack.track_ID], version, released_title, final_length)

Since these are some, but not all the tables in the *Orinoco* database, some of the overall semantics is missing, but broadly speaking, the database as a whole is intended to meet, among other, the following requirements:

- Each artist has an artistic name, works within a musical genre and is managed by a manager who has an ID, a name and can have several contact telephone numbers.
- Artists also have a contract with *Orinoco*, for which the following information is recorded: the period of the contract in terms of the date from which, as well as the date up to which, the contract holds. The duration of the contract in days is also stored.
- An artist records a number of master tracks, each of which has an ID, a working title, a working length. Information on the artist who recorded it and the sound engineer who edited it are also kept.
- A finished track originates from a master track and has a version number (as there may be different versions of it), a released title and a final length.
- A sound engineer has an ID and a name.
- An album groups together finished tracks into a sequence (i.e., a play list). An album has an ID, a title. Information is kept about the artist who created it. In this initial model, we are not including information that links an album with the finished tracks it contains. We will do that in later exercises, when we will also store information as to whether an album can be distributed as a CD, on vinyl, or as a cassette tape.

Requirements

From a machine running Linux, you will use Oracle SQL*Plus to interact with the Oracle DBMS. You will also use the Dia diagramming tool, and LaTeX which are also available in Linux machines.

Instructions

Firstly, work your way through Part 1, i.e., the SQL*Plus tutorial below.

Note what you need to submit in your report and learn how to save the information as you go. There are suggestions for that in the tutorial itself. The overall goal is for you to learn about the tables that you own, their definitions (i.e., the columns, and their types, that characterise each table) and the data that is stored in them (i.e., the rows that each table currently contains). You will also learn how to find out about the primary, and the foreign keys (along with their target primary keys) for the tables you own. Another important goal is for you to learn how to control the way SQL*Plus displays, and prints out to file, the results of command execution.

At the end of Part 1, i.e., the tutorial, you will be in a position to provide answers to items 1 to 3 below.

Secondly, in Part 2, use Dia to replicate the diagram shown in Figure 6 below. This will allow you to provide an answer to item 4 below.

Submission Contents

Your submission must contain:

- | | |
|--|-----------|
| 1. A listing of the structure of <u>any</u> 3 tables in the Orinoco database | [3 marks] |
| 2. A listing of the primary and foreign keys of <u>all</u> tables | [3 marks] |
| 3. A listing of the data in <u>any</u> 3 tables | [6 marks] |
| 4. A Dia diagram as described below | [6 marks] |

Additional marks are available as follows:

- | | |
|---|-----------|
| 5. Good report presentation (see the Lab Sessions/Example Clinics section above) | [2 marks] |
|---|-----------|

Part 1: SQL*Plus TUTORIAL

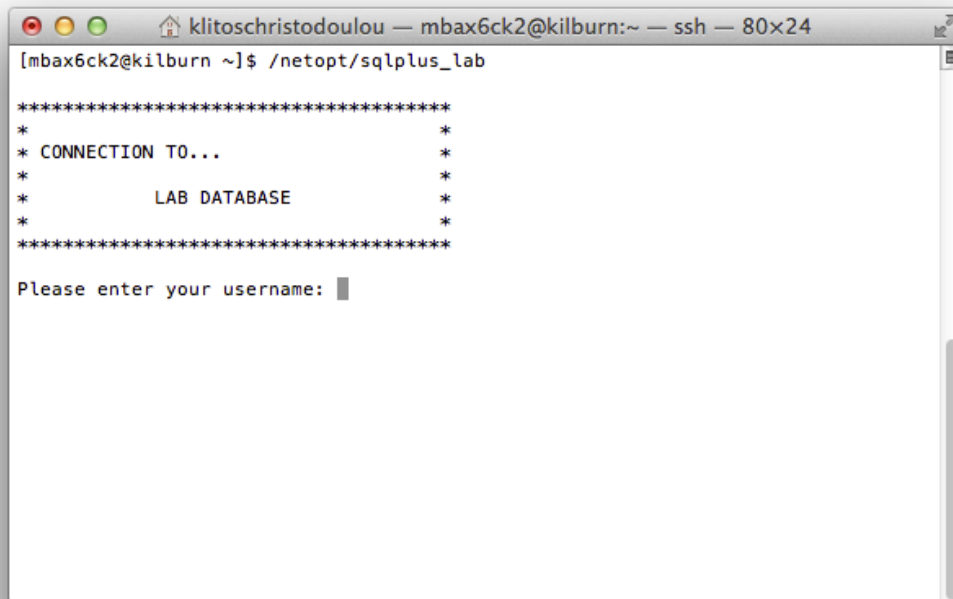
Starting SQL*Plus

Full information with examples for SQL*PLUS commands can be found in the online [Oracle documentation](#). In particular, the [SQL*Plus User's Guide and Reference](#) is also online.

You may also benefit from the e-books on SQL available online through the The University of Manchester Library (TUoML). Two that are useful for you to know about and consult are:

- **Beginning Oracle SQL**, Haan, L., Markham, J., Gorman, T., Pohlmann, F., Fink, D., Moodie, M., Buckingham, E., Apress, 2010. [Online Version](#)
- **Pro Oracle SQL**, Morton, K., Still, J., Shamsudeen, R., Osborne, K., and Sands, R., Apress, 2011. [Online Version](#).

For the lab session it is recommended that you use Linux. Follow the example in Figure LX-01 to launch SQL*Plus. You should use the user name and password sent to you. If you find a problem with the path even though you are running a Linux shell in a SCS UG lab machine, let the lecturers or TAs know. (If the full path does not work, try typing just sqlplus_lab.)



```

[mbax6ck2@kilburn ~]$ /netopt/sqlplus_lab

*****
*                                     *
*  CONNECTION TO...                  *
*                                     *
*      LAB DATABASE                  *
*                                     *
*****

Please enter your username: 
```

Figure LX-01: Logging in to Oracle SQL*Plus

When you see the SQL*Plus prompt, you can simply type in a SQL*Plus command. (Try `HELP TOPIC`, and you will have something to explore.)

We will aim to be precise in differentiating SQL*Plus *commands* (which are specific to Oracle) and SQL *statements* (which are, by and large, standard across many DBMSs). Also, we often refer to a SQL *query*, and ideally this would only apply to SQL statements that do not cause the database to change, but the use of the more precise word as a synonym for statement is very widespread and we'll probably fail to be consistent in our usage.

You can also paste commands/statements after typing and copying them in a text editor. (This is **highly recommended**, because retyping is tedious and SQL*Plus is not always as usable as, say, a modern Unix shell. For example, it doesn't support completion.)

Separate words with spaces or tabs (choose one approach and use it consistently). Commands and statements can appear in upper or lower case (again, choose one approach and use it consistently), but literals (as we will see) are case-sensitive.

There are also special SQL*Plus commands for formatting, setting options and editing and storing SQL commands (see the documentation for details and explore with the `HELP` command.). It is highly recommended that you explore.

Spooling Work

With the `SPPOOL <filename>` command, you can log all commands/statements plus their result into the text file `<filename>`. Use the `LabExNOut.lst` template, and since this is Lab Exercise 1, the filename becomes `LabEx1Out.lst`. When you want to stop logging, use `SPPOOL OFF`.

Running Scripts

You can run a textual sequence of SQL statements stored in a file as a script (using the `*.sql` suffix). In SQL*Plus, the `START <filename>` command executes the SQL commands in `<filename>`.

We have provided you with three files with SQL commands to give you something to play with. Their full paths are:

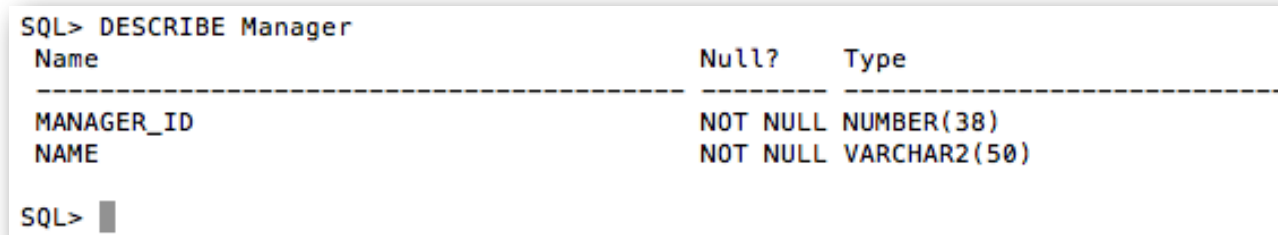
```
/opt/info/courses/COMP23111/create-Orinoco-initial-tables.sql
/opt/info/courses/COMP23111/drop-Orinoco-initial-tables.sql
/opt/info/courses/COMP23111/recreate-Orinoco-initial-tables.sql
```

The first one creates the tables mentioned above in the **Introduction** section and populates them with some data. The second deletes those tables (and their content). The third is essentially the append of the second to the first, and therefore returns the tables to the initial state by first deleting them, then recreating and repopulating them from scratch.

Use the SQL*Plus `START` command to execute the commands in the files above as you need. The first step is, of course, to create the tables. This produces lots of output. Just browse through it to make sure that no errors occurred. If so, then continue exploring.

Listing a Table Definition

One of the tables you now own is `Manager`. (Later in this tutorial, you will learn how to find all the tables you own.) Use the SQL*Plus command `DESCRIBE` to find out the definition of `Manager`, as it is held in the Oracle Data Dictionary. You should see what is shown in Figure LX-02, i.e., the column names and the types of the values stored (plus an indication as to whether the columns can store null values).



Name	Null?	Type
MANAGER_ID	NOT NULL	NUMBER(38)
NAME	NOT NULL	VARCHAR2(50)

Figure LX-02: Example Description of a Table

Make a mental note of the fact that SQL*Plus is printing lots of white space between value columns, the reason being that, in the data dictionary, the column where the `Name` attribute is stored is defined to be quite wide. You will learn to control what is printed, and you will need to exercise judgement as to readability and good layout in your lab exercise submissions if you are to merit the full marks for good report presentation. So, be attentive and make time to explore and learn on your own.

Ending a SQL statement

You generally will use a semicolon (`;`) to end a SQL statement. Hitting the enter/return key causes the statement to be executed. A statement can span many lines: so, it is the semicolon that denotes its end and it is needed, therefore.

Halting the execution of a SQL statement

If execution hangs (which should be very very very rare), use `Ctrl-C` to try and halt an execution.

Selecting all the rows and all the columns in a relation.

To select all columns and all rows in a table, instantiate the SQL template

```
SELECT * FROM <table_name>;
```

The '*' stands for 'all columns'. Try it with `Manager` as the table name.

Selecting some of columns from some of the rows.

To select some of the columns only, you list their names in the `SELECT` clause.

You can also select some of the rows only. If you also add a `WHERE` clause that specifies a predicate (i.e., a Boolean-valued expression) then a tuple must satisfy that predicate in order to be part of the output.

The full template is

```
SELECT <col_name1>, ..., <col_name_n>
FROM   <table_name>
WHERE  <predicate>;
```

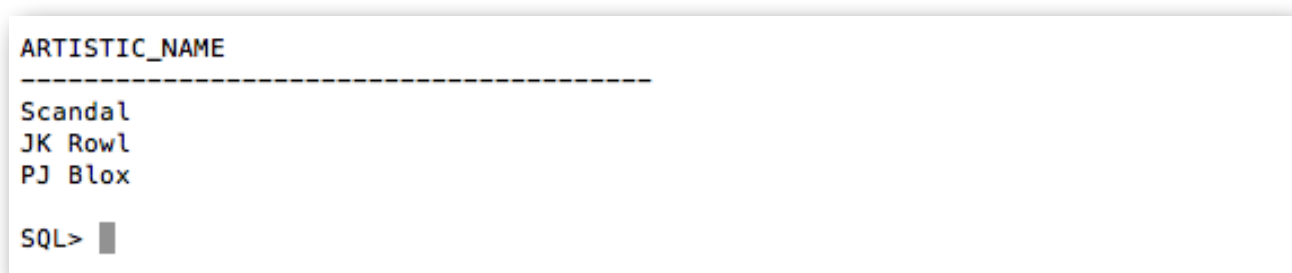
Try finding out the columns that comprise the `Artist` table. Then look at the entire content of the table.

Now, try finding just the names of the artists that satisfy the following predicate `Genre='Dance'`.

The single quotes to enter a string literal are important. Note also that while SQL and SQL*Plus are not, in general, case-sensitive, when you're passing literals in a statement (such as `'Dance'`) case does matter.

When using quotation marks for literals, be attentive to situations in which your text editor is too clever and generates an illegal character instead. Also, be careful when you copy and paste from PDF files; sometimes the character that gets pasted is not what it looks like.

When you've tried the above, this is the result you should see what is shown in Figure LX-03.



```
ARTISTIC_NAME
-----
Scandal
JK Rowl
PJ Blox

SQL> █
```

Figure LX-03: Example Output of a `SELECT-FROM-WHERE` query

After displaying the results and (often, but not always) the number of rows selected, SQL*Plus displays the command prompt again.

Ordering the results

You can use the additional template clause (after the WHERE clause, or the FROM clause if no WHERE clause is given)

```
ORDER BY <col_name1>, ..., <col_namen>
```

to sort the returned tuples in ascending, cascading order. Try sorting the results of your previous SQL statement by artistic name. Use the same statement as above but now add the appropriate ORDER BY clause.

Removing duplicates

Sometimes the results contain duplicates, but you can eliminate them, if you wish, by adding the keyword DISTINCT immediately after the SELECT keyword.

Creating and modifying SQL statements

Typically you will enter an SQL statement across several lines, execute it and then want to alter it (as we all make mistakes, and more so when we are learning) and try again.

It is highly recommended that you create and modify your statements using a text editor (avoid word processors, as they might tamper with, e.g., quote characters).

Type your statements into the text editor, then copy and-paste them into the SQL*Plus prompt. The line numbers may flash up and make the query look a bit of a mess, but they do not affect its execution. In this way, you can also maintain a history of different queries you have tried and edit/retry complex queries until you get the syntax right.

The most common problems are likely to be errors in the syntax that are difficult to identify from the error messages. You are strongly advised to break your queries into separate lines, and build a complex SQL statement up slowly, bit by bit. This will also help you gain a feel for compositionality, i.e., how a query, for example, can be made of subqueries (i.e., nested queries).

You can also make use of the arrow keys² to scroll up and down the command/statement history, but although this is often quicker, it is not as good for documenting your work as using the editor, so, use both methods and get the best of both worlds.

Dealing with errors

SQL*Plus has syntax rules for commands and so does SQL for statements in the language. If you break a syntax rule, you will get an error and the command/statement will not execute.

Most error messages will, in the usual way, give you a clue as to what you may have done wrong. If you get an error message that you cannot understand, learn how to use the online Oracle documentation (see the link at the start of this specification). If this is not helpful, ask the lecturers or TAs for help.

Try, for example, misspelling Manager as MInager on purpose and ask SQL*Plus to DESCRIBE it. You'll get an error. Then, in a web browser, use the search field in the online Oracle documentation webpage (see the link above) to try and find out more about the error by typing in the error code you got. Give or take some inessential differences, you should find in the documentation an explanation like the one shown in Figure LX-04.

² This is not standard SQL*Plus functionality. The server we use has been enhanced with this capability.

ORA-04043: object *string* does not exist

Cause: An object name was specified that was not recognized by the system. There are several possible causes:

- An invalid name for a table, view, sequence, procedure, function, package, or package body was entered. Since the system could not recognize the invalid name, it responded with the message that the named object does not exist.
- An attempt was made to rename an index or a cluster, or some other object that cannot be renamed.

Action: Check the spelling of the named object and rerun the code. (Valid names of tables, views, functions, etc. can be listed by querying the data dictionary.)

Figure LX-04: Example Explanation of an Error Message in Oracle's Online Documentation

Saving your queries, and the results of your queries, for inclusion in reports

For producing the final PDF document required for submission, remember to use the `LaTeX` template provided. For solutions to the tasks from each lab exercise you must incrementally create an SQL script file(s) in a text editor (for example, `ned`, or `vi`, or `vim`, or `emacs`), run the script(s) and, when you're satisfied with it, run it complete, making sure that you spool the result(s) into a separate `*.lst` file(s) for inclusion in your report.

Instructions on how to include the results into a `LaTeX` document are provided in the `ReadMe.txt` file located at:

```
opt/info/courses/COMP23111/lab_submission_N.zip
```

You should make a copy of the `lab_submission_N.zip` file into a local directory and `unzip` it using the following command:

```
$ unzip lab_submission_N.zip
```

then rename to appropriately the 'N' substring (as well as replacing, inside the folder, the place holders with your last and first name) in the names we provide.

Getting the results to look better formatted

When the results are returned in a format that makes them difficult for you to even read them, you must use the various commands that SQL*Plus offers to adjust display characteristics. To do so, explore them with `HELP INDEX`.

Use `HELP` to learn about the various things you can `SET`. Also look at the `COLUMN` command. Use `HELP` to learn how you can specify how columns get displayed, and so on. Finally, learn about the `CLEAR` command too. If needed, go beyond the help facility in SQL*Plus and search the Oracle online documentation.

Sensible use of the above can mean the difference between your reports being readable or not, and therefore meriting the good presentation marks or not.

The Oracle Data Dictionary

Like most DBMSs, Oracle has a data dictionary containing all the management information on the objects (i.e., tables, views, indexes, etc) in the database. You are not given authorisation to alter this dictionary directly, but you can look at it: this information is held in the form of tables too, so the usual SQL statements also work for the data dictionary tables.

The Oracle Database Administrator's Guide (see the online documentation) has copious amounts of information on that, as you would expect.

Listing the tables you own

The data dictionary table (well, it is actually a view) `user_tables` contains information on all the tables the user (i.e., you) own.

You can try asking SQL*Plus to DESCRIBE it, but you'll find that a lot of very technical information about the data dictionary comes flooding out. For the moment, the most interesting column is `table_name`. Since `user_tables` is just a table, you can access the information stored in it using SQL statements. For example, try this SQL statement:

```
SELECT table_name FROM user_tables;
```

Notice that the data strings in the dictionary are in upper case, so any WHERE clause that uses literals will have to be aware of this. This is so irrespective of what case was used in the strings that appeared in the SQL statement that created, altered or populated the database.

Use DESCRIBE to have a look at `user_tables` and see just how much metadata (i.e., data about data) an Oracle table has.

Integrity constraints in the data dictionary

The DESCRIBE command accesses information in the Oracle data dictionary. Notice that this does not tell us anything about keys or any integrity constraints defined on the table. We have to actually query the data dictionary to find out such information.

To find out the integrity constraints on, say, the `Manager` and the `ManPhone`, we need to look at two dictionary views:

- `user_constraints` for constraints on the tables
- `user_cons_columns` for constraints on the columns in the tables.

Starting with the tables

```
SELECT constraint_name, constraint_type,
       table_name, r_constraint_name,
       delete_rule, status
FROM   user_constraints
WHERE  table_name = 'MANAGER' or table_name = 'MANPHONE';
```

You may find that it prints something difficult to read. This is the time to set a wider line width. Find out how to use the SET command to extend the line width to 250, say, and then enlarge your window enough to accommodate a line as wide as that. Then, try again. You should see something like shown in Figure LX-05.

CONSTRAINT_NAME	C	TABLE_NAME	R_CONSTRAINT_NAME	DELETE_RULE	STATUS
SYS_C00250434	C	MANAGER			ENABLED
M_ID_PK	P	MANAGER			ENABLED
MP_ID_PK	P	MANPHONE			ENABLED
MP_ID_FK	R	MANPHONE	M_ID_PK	NO ACTION	ENABLED

SQL> █

Figure LX-05: Example Constraints on Tables

We need to go through this result. First of all, notice that the literal values in the dictionary table are all in upper case. This means that when they are queried for, they must be passed as upper case literals.

The constraint type (i.e., the values in the the column named C) can be:

- P Primary key, i.e., the identity integrity constraint
- R Foreign key, i.e., the referential integrity constraint
- C Constraint, i.e., a domain integrity constraint in the form of an SQL expression such as NOT NULL or CHECK IN ('cis','cs')

Constraints named SYS_XXXXXXX are created by Oracle for unnamed constraints (and will, most likely, be different in your case from what you see in Figure LX-05). Rows 2 and 3 state that the Manager and ManPhone have a primary key defined and that it is enabled. The constraints have been named by the owner (well, the script we gave you did that) with specific constraint names. Row 4 states that ManPhone has a foreign key pointing to the primary key of Manager.

The admissible values in the DELETE_RULE column for referential integrity constraints can be NO ACTION or CASCADE (which we will learn about in due course). Finally, the STATUS of the constraint can be ENABLED or not.

Now, we can find all type C constraints using:

```
SELECT constraint_name, table_name, search_condition
FROM   user_constraints
WHERE  constraint_type = 'C' AND table_name NOT LIKE '%BIN%';
```

(Disregard, for now, the condition table_name NOT LIKE '%BIN%'. It is being used here to filter out the many constraints that Oracle generates of its own accord. If you are curious and want to have a look, just rerun the query without the conjunct that filters it. You can ignore the system generated information: the data dictionary is a repository of both user-level and system-level information, so querying it can be, as here, a bit messy.)

The tuples in the result all have, of course, a C-type constraint. The C stands for 'check'. In this case, to check that particular attributes are NOT NULL.

The user_constraints table/view just tells us on which tables there are constraints, but not the columns that the constraints act upon. To see this we have to look at the user_cons_columns table/view.

Use the following query now:

```
SELECT constraint_name, table_name, column_name, position
FROM   user_cons_columns
WHERE  table_name = 'MANAGER' or table_name = 'MANPHONE';
```

Again, this will likely print in a difficult-to-read way. You will find that the column COLUMN_NAME is too wide. It is time to tinker with the way it is displayed. Recall that we suggested that you looked into what the COLUMN command in SQL*Plus does. You should have learned to how use it, so set the column COLUMN_NAME to a narrower width, as follows:

```
COLUMN column_name FORMAT a20
```

and try again.

We can now see that the constraint MP_ID_PK refers to column Telephone in the table ManPhone. The name suggests that this is a primary key constraint. It also tells us that MP_ID_FK in the ManPhone table refers to the column with name MANAGER_ID, suggesting that this is a foreign key constraint referring to the Manager table's primary key (identified by MP_ID_PK). We can infer much of this information from the

constraint names if we always make an effort to give constraints sensible names. (Note that POSITION refers to the position of a column in a composite key.)

Reloading the tables if they have become corrupted

If you've managed to delete or otherwise damage the data in the tables, use the recreation script mentioned above.

This is the end of the tutorial, though you are expected to explore much much further on your own, at your own time.

For now, work on items 1-3 of your submission. Remember to save and organise your input and output and to incorporate it to the report you must submit. Then, move on to Part 2.

Part 2: Getting Started with Dia

[Dia](#) is a Linux application for creating technical diagrams, whose interface and features are loosely based on Visio, the Microsoft Windows program.

Dia is easy enough to learn without much hassle and flexible enough to make power users feel at home when compared to commercial tools.

The first thing to do with Dia is to have a look at the [online Dia documentation](#). Then, play with the tool and familiarise yourself with it.

Now, consider Figure LX-06. It gives you an indication of the notation for (E)ER modelling that Dia supports.

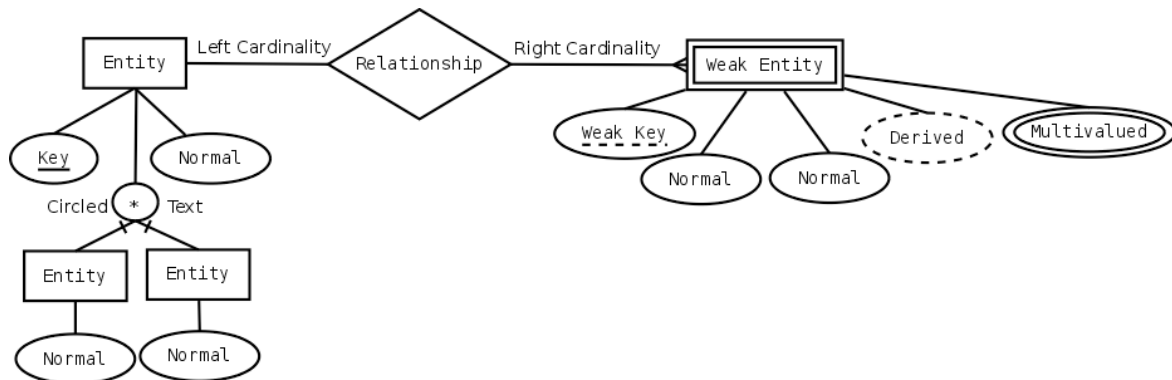


Figure LX-06: Legend: (E)ER Constructs in Dia

We have not covered (E)ER modelling yet, but we soon will. For this lab exercise, you do not need to worry about what each of the symbols, notation and concepts actually mean in detail. All you need to do is to use Figure LX-06 as a crib (by structural similarity) for you to reproduce Figure LX-07.

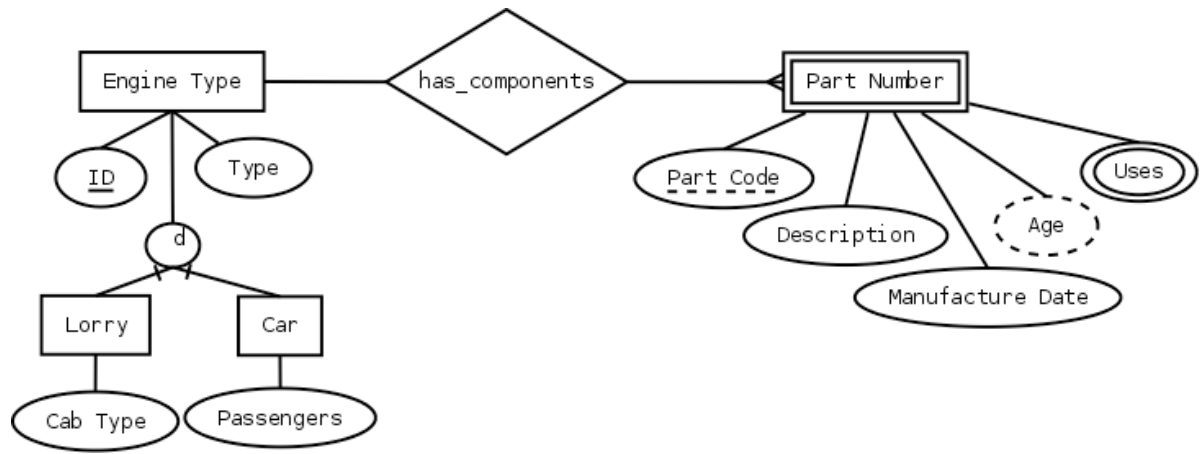


Figure LX-07: An Auto Manufacturer (E)ER Diagram in Dia

In other words, your objective is to use Dia and, using Figure LX-06 as your guide, create the diagram in Figure LX-07 in Dia, which is what will merit marks. You do not need to draw both diagrams, you only need to draw the one in Figure LX-07. Also, there is nothing tricky going on here: all we are asking you to do is, really, to replicate the diagram using Dia. We just want to make sure that you have used the tool and learned the basic functionality needed.

Work out by yourself how to draw the bottom part by exploring the tool and referring to the online documentation if you need to. A hint here is for you to, first of all, learn (either exploring or reading the online documentation) how Dia supports different diagrammatic notations and, in particular, how it supports the drawing of (E)ER diagrams. Another hint is that some lines are special, but do not look so. A final hint is that not all constructs you see drawn are supported in the same way (e.g., in terms of entering text into a shape).

Once you have finished, make sure you save your diagram as a .dia file in the appropriate directory (as with all other parts of any submission). Also, using Dia, you need to export (File -> export) the diagram into an EPS figure and include it in the LaTeX document. Do study the `ReadMe.txt` file for detailed instructions.

You are required to submit all the files used to generate the report using the LaTeX template provided. This includes the .dia file, any .lst file(s) and the SQL script(s) you created, all stored in their appropriate directories. Again, see the `ReadMe.txt` file.

Examples Clinic 2

Entity-Relationship Modelling

Specification

The goal of this examples clinic is to provide you with some practice on entity-relationship (ER) modelling. It aims to prepare you for the lab exercise in which you are asked to explain and extend an ER model for the Orinoco database that was used in the first lab exercise.

DURATION: 50 minutes examples clinic time.

DESCRIPTION: The examples clinic consists of two tasks: the first one asks you to draw an ER diagram given the specification of the data requirements; the second one asks you to reverse engineer the specification of the data requirements given an ER diagram. Work through the questions, demonstrate them to the lecturers and TAs in order to obtain feedback.

TASKS

T1. Consider the following set of requirements for an American UNIVERSITY database that is used to keep track of students' transcripts:

- I. The university keeps track of each student's name, student number, social security number, current address and phone, permanent address and phone, birthdate, sex, class (freshman, sophomore, ..., graduate), major department, minor department (if any), and degree program (BSc, MSc., ..., PhD). Some user applications need to refer to the city, state, and zip of the student's permanent address, and to the student's last name. Both social security number and student number have unique values for each student.
- II. Each department is described by a name, department code, office number, office phone, and college. Both name and code have unique values for each department.
- III. Each course has a course name, description, course number, number of semester hours, level, and offering department. The value of course number is unique for each course.
- IV. Each section has an instructor, semester, year, course, and 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.
- V. A grade report has a student, section, letter grade, and numeric grade (0, 1, 2, 3, 4 and F, D, C, B, A, respectively).

Design an ER schema for this application, and draw an ER diagram for that schema. Specify the key attributes of each entity type and the structural constraints on each relationship type. Note any unspecified requirements, and make appropriate assumptions to make the specification complete.

T2. Consider the ER diagram in Figure EC-02, which shows a simplified schema for an airline reservations system. Reverse-engineer from the ER diagram the textual specification of requirements and constraints that resulted in this schema. Try to be as precise as possible.

Figure 7.20

An ER diagram for an AIRLINE database schema.

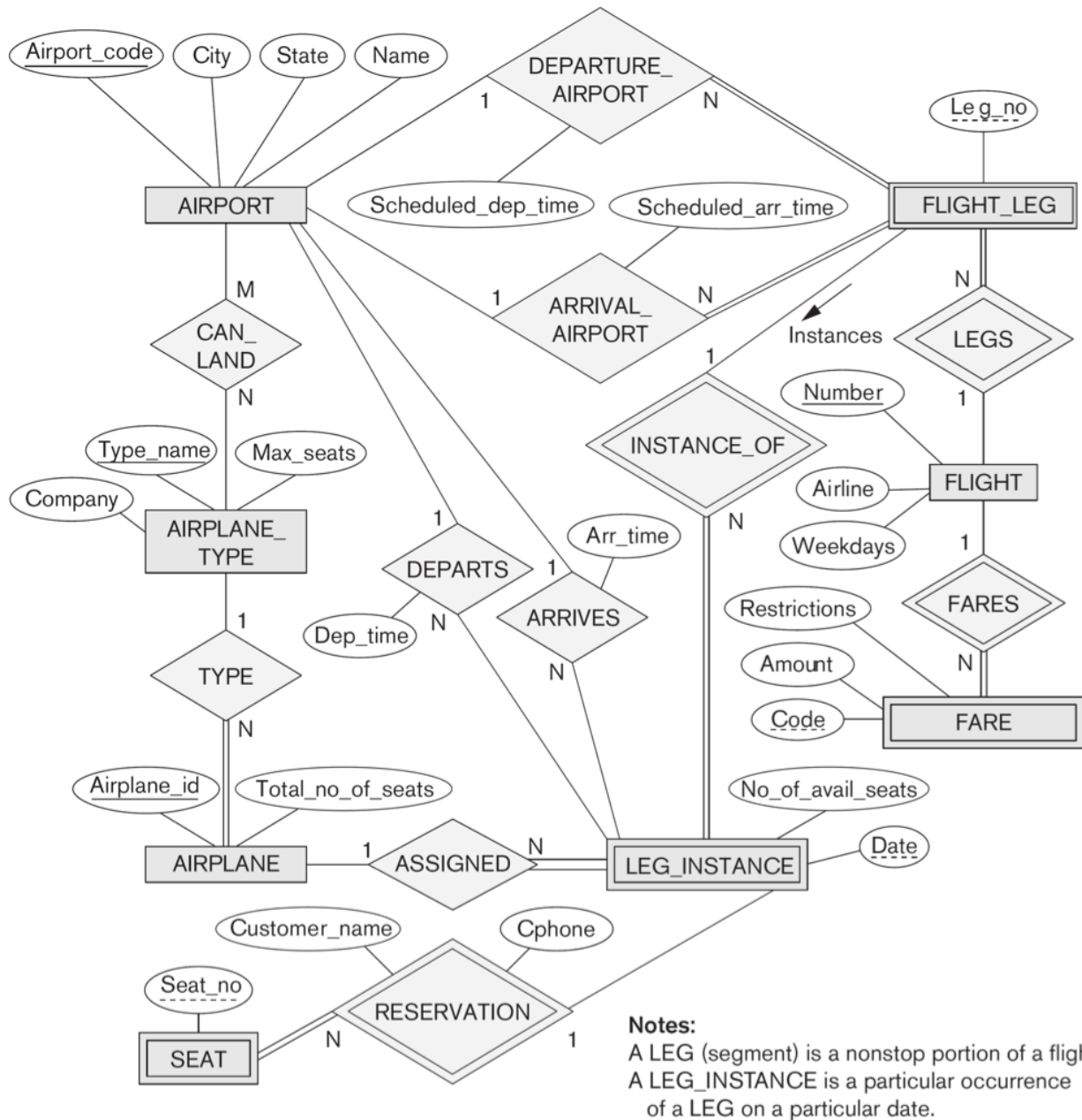


Figure EC-02: ER Diagram for an Airline Reservations System

Acknowledgements

The tasks in this practical session are minor adaptations of material authored and made available to instructors by Ramez Elmasri and Shamkant B. Navathe to accompany their textbook Database Systems: Models, Languages, Design, and Application Programming, 6th (Global) Edition, Addison-Wesley Pearson, 2011, 978-0-13-214498-8 Copyright © 2011 Pearson Education, Inc., whom we thank. All errors are our responsibility.

Lab Exercise 02

(E)ER Modelling Specification

Overview

The goal of this lab exercise is to provide you with an opportunity to gain some practice in the conceptual design of databases.

You will start with a conceptual schema in the form of a ER diagram that represents the *Orinoco* case-study database that you used in Lab Exercise 1. You will be asked to write an explanation of concepts in the given diagram to demonstrate that you understand what information an ER diagram captures. We will then provide you with additional requirements for the *Orinoco* database and you will proceed to extend the given ER diagram to an extended (E)ER diagram so as to capture the new requirements.

ATTENDANCE:	Compulsory
MARKS AVAILABLE:	20 marks.
DEMONSTRATION:	You are required to submit two files: a report in PDF format generated from the template we provide and an archived folder of the outcomes of your work (scripts, spooling output, and diagrams).
TIME AVAILABLE:	2 hour lab session.
DEADLINE:	The end of the session.
EXTENSIONS:	Provided you attended the lab session and worked on the exercise all the time, you are granted an automatic extension. No further extensions are granted without documented special circumstances. This is strictly enforced.

Introduction

This lab exercise has two parts: in **Part 1**, your aim is to explain concepts (and in one case write down the data requirements) that can be inferred from a given ER diagram; in **Part 2**, your aim is to extend the given ER diagram to an enhanced entity-relationship (E)ER diagram that captures the additional data requirements provided.

Figure LX-08 (below) shows an ER schema corresponding to the (partial) *Orinoco* database as created in Lab Exercise 1 in Oracle using SQL scripts. This conceptual schema will be used as a starting point for this exercise and future laboratory exercises as well. The corresponding .dia file is available here

`/opt/info/courses/COMP23111/OrinocoInitial.dia`

you should make a copy of it in whatever directory in your personal file space you are going to work from in this lab exercise.

Requirements

From a Linux machine, you will use the Dia diagramming tool. Note that some diagrammatic conventions for constructs that belong to the Enhanced ER model may not be directly supported in Dia.

Instructions

Firstly, work your way through Part 1. Note what you need to submit and learn how to save the information as you go. At the end of the Part 1 you will be in a position to provide answers to items 1 to 6 below.

Secondly, in Part 2, use Dia to extend the diagram shown in Figure 1 below (working on the corresponding .dia file). This will allow you to provide an answer to item 7 below.

Finally, study the additional requirements and your extended diagram and provide the answers sought in items 8 to 10 below.

Submission Contents

Your submission must contain:

1. An answer to Item 1 in Part 1 [1 mark]
2. An answer to Item 2 in Part 1 [1 mark]
3. An answer to Item 3 in Part 1 [1 mark]
4. An answer to Item 4 in Part 1 [1 mark]
5. An answer to Item 5 in Part 1 [1 mark]
6. An answer to Parts a-b of Item 6 in Part 1 [4 marks]
7. A figure, i.e., a Dia-drawn (E)ER diagram (*), including the requirements in Item 7 in Part 2 [9 marks]

Additional marks are available as follows:

8. Good report presentation (see the **Lab Sessions/Example Clinics** section above) [2 marks]

Part 1: Understanding the ER schema for the partial Orinoco database

To demonstrate that you understand the information represented by the schema, consider Figure LX-08.

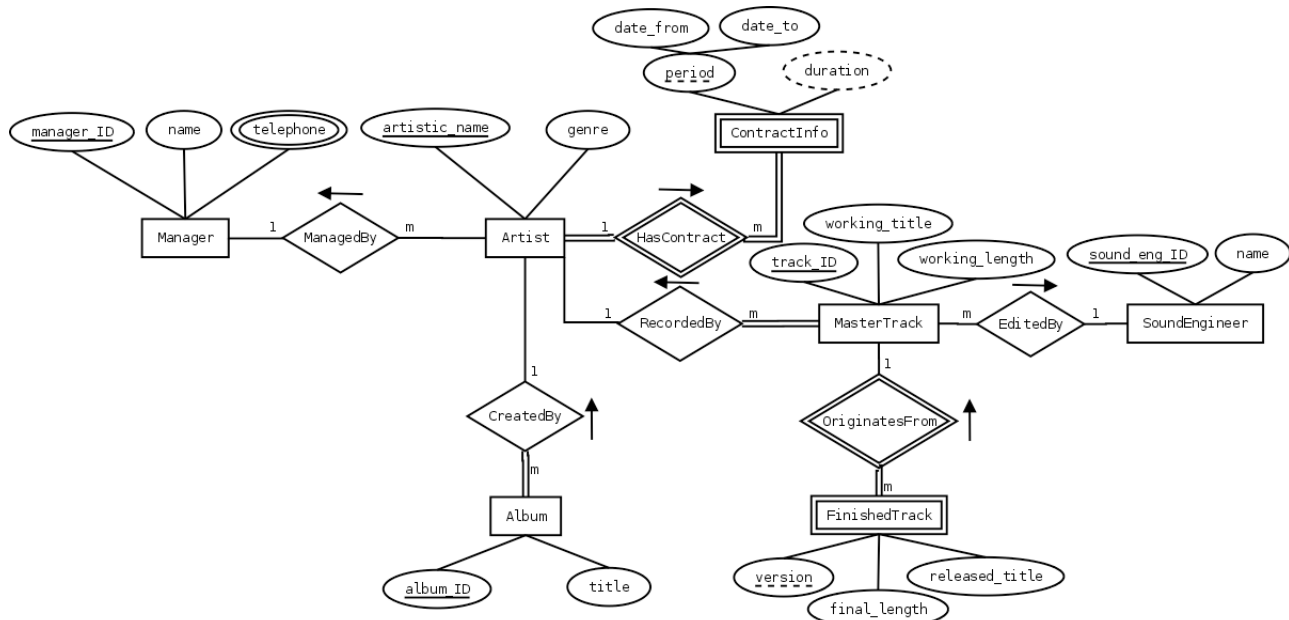


Figure LX-08: ER Diagram for the (Partial) Orinoco Database

Using Figure LX-08 (but you should also work with your copy of the corresponding .dia file in Dia), you are asked to explain the following:

1. What kind of attribute is the telephone attribute of the Manager entity type?
2. What kind of entity type is ContractInfo? What condition causes an entity type to be categorised as being of this kind?
3. What kind of relationship type is HasContract? What condition causes a relationship type to be categorised as being of this kind?
4. What kind of attribute is the period attribute of the ContractInfo entity type? What does the dashed underline underneath it mean? (Do not answer the second question with the same answer you gave to the first! We're looking for two pieces of knowledge here: what is it called and what is/are its defining characteristic/s!)
5. What kind of attribute is the duration attribute of the ContractInfo entity type? What does the dashed oval that encircles it mean? (Again, do not just repeat for the second question the answer you gave to the first! Again, we're looking for two pieces of knowledge here: what is it called and what is/are its defining characteristic/s!)

6. Write no more than three or four short paragraphs in the style of a data requirements specification document³ that might have led to the diagrammatic model you see in Figure 1 for:
 - a. the MasterTrack entity type and its attributes,
 - b. the three relationship types in which it participates including their corresponding cardinality ratio and participation constraints.

If you find you are writing more than half a page in giving answers to (1)-(6) above, you are probably misunderstanding what is being asked.

Part 2: Extending the ER schema for the Orinoco database

To demonstrate you can create ER conceptual schemas, you are now asked to extend the ER schema depicted in Figure LX-08 to an enhanced entity-relationship (E)ER diagram, which corresponds to the Orinoco database as you have created it in Lab Exercise 1 in Oracle using SQL scripts.

7. You should start with your copy of the given .dia file (see above for its path/location) and, using the Dia program, extend the diagram in order to capture the following new requirements:
 - a. Recall that in Lab Exercise 1 we said that, in the initial database we gave you to work with then, we did not include information that linked an album with the finished tracks it contains. It is now time for you to model the fact that an album must have at least one, and normally many, finished tracks and that a finished track may appear in many albums. You must also capture the sequence in which finished tracks appear in an album. Note that the same finished track may appear in different albums in different sequences.
 - b. The system must keep an online collection of catalogue entries, each of which references an existing album. An album must be listed as a catalogue entry at least once, and may be listed as several catalogue entries. Each catalogue entry is differentiated by its release date (to account for re-releases of the album). Both the price and the stock for each catalogue entry are also stored.
 - c. We record the name of buyers as well as the date at which they have registered with the Orinoco store. We also record a buyer ID to uniquely identify buyers.
 - d. A buyer can place any number of orders. An order is identified by an order number. We also store the date an order has been placed on by a buyer and the dispatched date of an order. An order comprises several catalogue entries. A catalogue entry can be part of many orders.
 - e. We want to model the fact that there exist two different types of artists: solo artists and group artists. A group artist can have as members more than one solo artist. You must also capture the date a solo artist has joined a particular group and the fact that a solo artist can be a member of more than one group. For solo artists, we would also like to store their real names and when they had their first performance. For group artists we would like to store the date a particular group has been formed. You should consider whether this specialisation leads to overlapping or disjoint sub-entities.
 - f. We also want to model the fact that an album is distributed as a vinyl album and/or a tape album and/or a CD album. A vinyl album can have many distinct colours, a CD album can have many different extras (e.g., an associated videoclip), and a tape can have exactly one label. You should also consider whether this specialisation leads to overlapping or disjoint sub-entities. (Hint: When a book is released in several forms, e.g., hardback, paperback, audiobook, does each format merit a different ISBN or not? So, is a book in hardback the same entity as a book in paperback or not? So, are hardback books and paperback books overlapping or disjoint sub-entity types?)

Now, you should think about the following (optional)⁴:

8. Is it possible for many artists to appear on one album? (Hint: follow the relationship paths and keep the cardinality ratios in mind.)

³ For an example of what is meant here by a *data requirements specification document*, see Chapter 7, Section 7.2, p. 197-198 of Elmasri And Navathe's Database Systems (6th ed.). This chapter is part of the collection of PDF files made available to you in the Mandatory Readings assignments. See also item 7 in this lab exercise for another example of what form such specifications can take. You do not need to follow any style strictly but you do need to convey in words the relevant information that is captured in that required parts of the ER diagram.

⁴ Things to think about. There are no marks associated with this task.

Examples Clinic 3

Mapping Entity-Relationship Diagrams into Relational Schemas Specification

The goal of this examples clinic is to provide you with some practice on mapping entity-relationship (ER) conceptual models in the form of an ER diagram into a relational schema. It aims to prepare you for the lab exercise in which you are asked to map the ER model for the *Orinoco* database into a relational database schema.

DURATION: 50 minutes examples clinic time.

DESCRIPTION: The examples clinic consists of two tasks: the first one asks you to map a given ER schema (illustrated in Figure 1) into a relational schema; the second one asks you to specify ALL primary and foreign keys for the schema you generated for the first task. Work through the questions, demonstrate them to the lecturers and TAs in order to obtain feedback.

TASKS

T1. Consider the ER schema in Figure 1 for a database that may be used to keep track of transport ships and their locations for maritime authorities. Map this schema into a relational schema. Use the seven-step algorithm presented in the lectures.

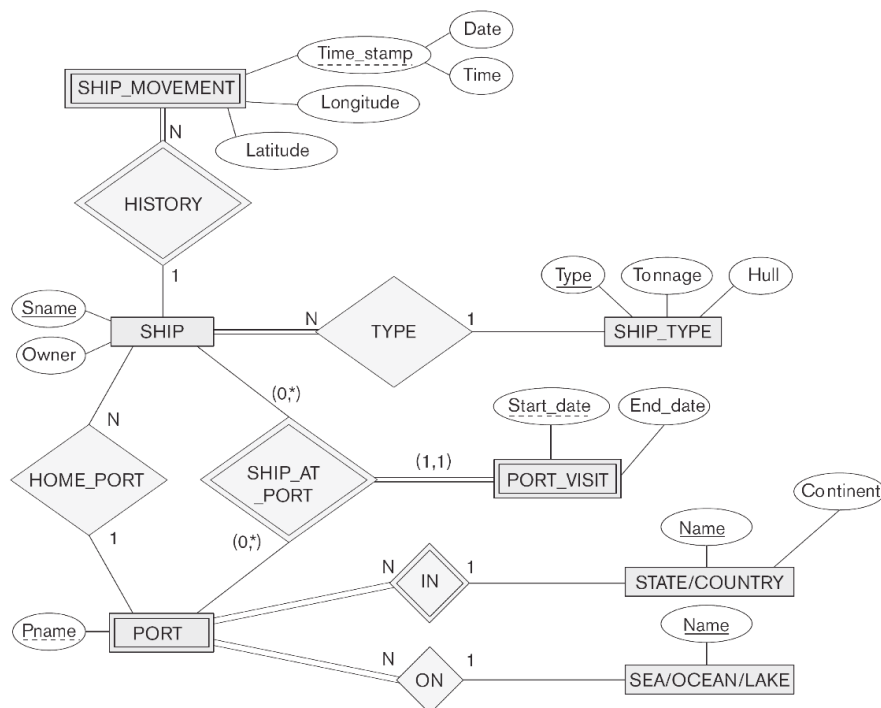


Figure EC-03: ER Diagram for a Maritime Transport Tracking Database

T2. Specify all primary keys and foreign keys for the schema you generated in **Q1**.

Acknowledgements

The tasks in this practical session are minor adaptations of material authored and made available to instructors by Ramez Elmasri and Shamkant B. Navathe to accompany their textbook Database Systems: Models, Languages, Design, and Application Programming, 6th (Global) Edition, Addison-Wesley Pearson, 2011, 978-0-13-214498-8 Copyright © 2011 Pearson Education, Inc., whom we thank. Any errors are our responsibility. whom we thank.

Lab Exercise 03

EER-to-R Mapping, Functional Dependencies, SQL DDL/DML Specification

Overview

The goal of this lab exercise is to provide you with an opportunity to gain some practice in mapping a conceptual schema onto a relational one, and in identifying what functional dependencies hold in the latter, as well as in generating SQL DDL statements that populate a database with relations and the latter with instances.

You will start with your conceptual schema in the form of an (E)ER diagram that represents the *Orinoco* case-study database as it stands after the extensions carried out in Lab Exercise 2. You **must** use the EER diagram that you submitted for Lab Exercise 2 but you should incorporate the feedback provided to you. If you have doubts about this feedback and how to get your (E)ER diagram into a correct state, talk to the lecturers/TAs of the course.

We will provide you with a partial relational schema for the *Orinoco* database (almost corresponding to the ER diagram as it was at the start of Lab Exercise 2). You will be asked to complete this mapping task so that you obtain a relational schema for the *Orinoco* database as it stands after the extensions carried out in Lab Exercise 2. You will also be asked to identify the functional dependencies that hold in the relational schema.

You will also rely on the scripts used in Lab Exercise 1. You will be asked to extend them with SQL statements that create the new tables resulting from the extensions carried out in Lab Exercise 2 as well with SQL statements that populate the new tables.

ATTENDANCE:	Compulsory
MARKS AVAILABLE:	20 marks.
DEMONSTRATION:	You are required to submit two files: a report in PDF format generated from the template we provide and an archived folder of the outcomes of your work (scripts, spooling output, and diagrams).
TIME AVAILABLE:	2 hour lab session.
DEADLINE:	The end of the session.
EXTENSIONS:	Provided you attended the lab session and worked on the exercise all the time, you are granted an automatic extension. No further extensions are granted without documented special circumstances. This is strictly enforced.

Introduction

This lab exercise has two parts: in **Part 1**, your main aim is to complete the EER-to-relational mapping task so as to capture the extensions to the *Orinoco* database carried out in Lab Exercise 2; in **Part 2**, your main aim is to extend and execute the SQL scripts used in Lab Exercise 1 to cater for the extensions made in the *Orinoco* database since Lab Exercise 1.

Figure LX-09 (below) shows a partial relational schema, capturing the ER diagram of the *Orinoco* database (in a textual form) as it was at the start of Lab Exercise 2.

Requirements

From a Linux machine, you will use SQL*Plus to access the Oracle DBMS. The instructions for that were given in Lab Exercise 1. You should use the same user-name and password as you used in Lab Exercise 1.

Instructions

Firstly, work your way through Part 1. This should result in a complete relational schema for the extended *Orinoco* database that corresponds to the conceptual schema you submitted for Lab Exercise 2 and includes the partial relational schema in Figure LX-09. You are not required to explain the workings of the mapping algorithm, but you can do this if you wish and also, as always, make explicit any assumptions you feel you

need to make if you believe information is missing, incomplete or conflicting. Then, working from the original SQL scripts, create the new one(s) that meet the requirements in Part 2.

Submission Contents

Your submission must contain:

1. The complete relational schema for the extended *Orinoco* database resulting from Part 1 [8 marks]
2. The SQL script resulting from Part 2 [10 marks]

Additional marks are available as follows:

3. Good report presentation (see the **Lab Sessions/Example Clinics** section above) [2 marks]

Part 1: Applying the EER-to-relational mapping to the extended Orinoco database

To demonstrate that you know how to map a conceptual schema to a relational one using the algorithm described and discussed in the lectures and available in the handouts (and the assigned readings), take as your starting point the conceptual schema for the extended *Orinoco* database, as shown in a textual form in Figure LX-09 (below). This shows the *Orinoco* database as it stood at the end of Lab Exercise 2.

Figure LX-09 shows the outcome of the EER-to-relational mapping algorithm if applied to the original *Orinoco* database, represented by the ER diagram you were given at the start of Lab Exercise 2. This part of this lab is about completing the relational schema in Figure LX-09 (below) so that it captures all the information in the conceptual schema you have submitted for Lab Exercise 2.

Using Figure LX-09 as your starting point, you must do the following:

1. Complete the mapping from EER-to-relational but going through all the entity types and relationship types (as well as their attributes, of course) that were added after Lab Exercise 2 (as reflected in your submitted solution)
2. Make sure that you indicate clearly which are the (possibly composite) primary keys as well as the foreign keys (and, for the latter, what primary key in what relation the foreign key references).
3. For each of the relations you have added as a result of Steps 1-2 above, list the functional dependencies that they exhibit.

Manager	(<u>manager_ID</u> , name)
ManPhone	(<u>manager_ID</u> [fk:Manager.manager_ID], telephone)
Artist	(<u>artistic_name</u> , genre, managedBy [fk:Manager.manager_ID])
Album	(<u>album_ID</u> , title, createdBy [fk:Artist.artistic_name])
MasterTrack	(<u>track_ID</u> , working_title, working_length, recordedBy [fk:Artist.artistic_name], editedBy [fk:SoundEngineer.sound_eng_id])
SoundEngineer	(<u>sound_eng_ID</u> , name)
ContractInfo	(<u>hasContract</u> [fk:Artist.artistic_name], <u>date_from</u> , <u>date_to</u> , duration [derived as (date_to - date_from)])
FinishedTrack	(<u>originatesFrom</u> [fk:MasterTrack.track_ID], <u>version</u> , released_title, final_length)

Figure LX-09: Relational schema partially corresponding to the ER Diagram for the original *Orinoco* database

Part 2: Updating the schema/populating new relations in the Orinoco database

To demonstrate that you know how to use the basic data definition and data manipulation statement in SQL, take as your starting point the following:

- The relational schema you have created in Steps 1-2 in Part 1 above.

- The initial, original *Orinoco* database implemented in the Oracle DBMS in the state it was left at end of Lab Exercise 1. Remember that, in Lab Exercise 1, three SQL scripts were made available for you to create, delete and repair the tables and tuples in the original, initial *Orinoco* database.
- Consider again the scripts used in Lab Exercise 1 (you may have a personal copy in your own file space that you made when you did that lab exercise):

```
/opt/info/courses/COMP23111/create-Orinoco-initial-tables.sql
/opt/info/courses/COMP23111/drop-Orinoco-initial-tables.sql
/opt/info/courses/COMP23111/recreate-Orinoco-initial-tables.sql
```

The goal in this part of the course is for you to develop an SQL script (you should call it `create-Orinoco-extended-tables.sql`) that does for the extended *Orinoco* database the same job that the original SQL script did for the initial, original *Orinoco* database in Lab Exercise 1.

Now you must do the following:

1. Create the new script(s)⁵ making sure that it contains:
2. All the table-creation information contained in the original script, except that you must make the following adaptations:
 1. All the table-creation information for the new relations that you derived in Steps 1-3 above (in doing so, you are advised to follow the style and conventions of the original scripts);
 2. All the SQL statements need to populate the tables as follows:
 1. The instance-level information in the original SQL script must be replicated in the new script(s) for the following tables: `Manager`, `ManPhone`, `Artist`, `Album`, `MasterTrack`, `SoundEngineer`, `ContractInfo` and `FinishedTrack`.
 2. The following buyers have registered with Orinoco: 'Simon Plant' on 28th Oct 2009; 'Julie Last' on 04th Aug 2011; 'Rina Blue' on 10th Mar 2012; 'Kay Sraonian' on 20th Dec 2009; and 'Yos Willis' on 10th Jan 2008. Buyers were assigned IDs 1, 2, 3, 4, and 5, respectively.
 3. Of the artists already present in the `Artist` table some are solo artists and some are groups. Some of the solo artists are members of one or more groups. The following is a complete list:
 1. Simon Palm's real name is 'Fred Maynard': he's a solo artist performing since 09th of Jan 2004 but is also a member of both the groups 'Goldfrat' since 23rd of Nov 2010 and 'Scandal' since 12th of Nov 2012.
 2. John Cliff has no stage name, he is known just as 'John Cliff': he's a solo artist performing since 12th of Dec 2005 and a member of 'Goldfrat' since 10th of Oct 2010.
 3. Jay Blacncar's real name is 'Mervin James': he has been performing as a solo artist since 09th of Jun 2003 but is also a member of 'Goldfrat' since 10th of Oct 2010.
 4. Palmer John is really called 'Samantha Jules': she has been performing as a solo artist since 12th of Feb 2006 but is also a member of 'Flut' and 'Scandal' since 08th of Jun 2009 and 20th of Oct 2012 respectively.
 5. JZ is really called 'Julie Zeb': she's a solo artist performing since 09th of Mar 2004 but is also a member of both 'Flut' and 'Scandal' since 10th of Dec 2009 and 20th of Oct 2012 respectively.
 6. JK Row's real name is 'Marisa Jaher': she has been performing as a solo artist since 11th of Jan 2003 but is also a member of both 'Flut' since 08th of Jun 2009 and 'Zero7' since 15th of Dec 2011.
 7. Finally, PJ Blox's real name is 'Pandar Rahoux': he's a solo artist performing since 04th of Apr 2007 but is also a member 'Zero7' since 15th of Dec 2011.
 4. Additional information on albums is as follows:
 1. 'Goldfrat' have recorded one album called '20 People in a Field', which has been released on CD, on vinyl and on tape, with assigned IDs '1c', '1v' and '1t' respectively. There are 2 CD releases: one with extra information available in the 'Web', and another with an extra 'Video'. There are also 3 vinyl releases: one on 'Red' vinyl, another one on 'Black', and a third one on 'Green'. The tape release has a 'Fancy' label.

⁵ You may wish also to study the other two scripts, the one that deletes tables and the one that repairs them. Notice that the repair script is nothing more than the concatenation of the delete and the create scripts, in that specific order. Then, after you have done the script that creates the tables (including the new ones), it may make your life easier to recreate the delete and repair scripts for the new database. However, pay attention that because there are dependencies between tables you may need to propagate the deletion. Look up the search terms `DROP CASCADE` in the Oracle documentation. Notice, furthermore, that it may make your life easier to break the original create scripts into two separate scripts: one that script that creates the tables and indexes, and another script that contains the statements to populate the tables created by the previous script. This means that you'd have two scripts rather than just one. For your submission, you can concatenate them back into a single one again.

2. 'My Feet' by 'Flut' is released on CD (with an extra 'Game') and on 'Red' vinyl with assigned IDs '2c' and '2v' respectively.
3. Finally, Jay Blancard's album 'Debut' is only released on CD with assigned ID '3c' but has two versions: one with extra 'Video' and another with an extra 'Game'.
5. The information about album tracks is as follows:
 1. The CD version of '20 People in a Field' has the following tracks in the sequence they are listed here: 'Hello', 'Extended Hello', 'Me', 'You', 'Them', 'Us', 'How', 'Be', 'Peace' and 'Smile'.
 2. The vinyl and tape versions of '20 People in a Field' only have the first 7 tracks of the CD version.
 3. Both the CD and vinyl versions of 'My Feet' have the following tracks in the sequence they are listed here: 'Wow', 'Whizzy', 'Who Are You?', 'Place and Time', 'Milko', 'Mad', and 'Bling'.
 4. Finally, 'Debut' has the following tracks in the sequence they are listed here: 'Power', 'What is your name?', 'Mix', 'Manic', 'Streets', 'Road', 'Forest', 'Deep', 'Danger' and 'Danger Remix'.
6. The online catalogue has been populated with entries for sale to the buyers. Each of the albums has an entry:
 1. '20 People in a Field' was firstly released on 24th of Dec 2010 as a tape, on 10th of Jan 2011 as a vinyl and five days later, on 15th of Jan 2011 as a CD. It is priced at £14.99 on CD for which the current stock is 2, it is priced at £12.99 on vinyl for which the current stock is 5, and, finally, it is priced at £10.99 on tape, for which the current stock is 9 tapes.
 2. 'My Feet' was firstly released as a vinyl on 20th of Jan 2010 and then on 15th Nov 2010 as a CD album. It is priced at £14.99 on CD and there is 1 copy in stock, while on vinyl it is priced at £12.99 but there is no stock left.
 3. Finally, 'Debut' by Jay Blancard was released on 10th Dec 2008. It is available only as a CD album at a price of £14.99 and there is one copy left in stock available.
7. Three orders have been placed, as follows:
 1. The first order was placed by Simon Plant, on the 17th of Jan 2011 and was dispatched three days later on the 20th. It had been assigned order ID=1 and consisted of one catalogue entry: Goldfrat's '20 People in a Field' on CD.
 2. The second order was dispatched to Rina Blue on the 15th of Mar 2012, after it was placed on the 12th of Mar 2012. It had been assigned order ID=2 and was for two catalogue entries: Goldfrat's '20 People in a Field' but this time on tape and the vinyl album 'My Feet' by 'Flut'.
 3. Finally, the third order had been placed again by Simon Plant on the 06th of Dec 2011 and was dispatched on the 10th of Dec 2011. It had been assigned order ID=3 and was for three catalogue entries: Goldfrat's '20 People in a Field' but this time on vinyl, the CD album 'My Feet' by 'Flut' and the CD album 'Debut' by 'Jay Blancard'.
8. Indexes provide opportunities for fast retrieval. We'll study them later in the course. For this exercise, you'll just imitate the index creation in the original SQL scripts you were given in Lab Exercise 1. In general, most DBMSs tend to create indexes on primary keys and foreign keys by default. However, most database applications that interact with humans also benefit from fast retrieval on human-readable information. Primary key and foreign key indexes are typically not (or not easily) human readable. You can see in the original SQL script that some indexes were created that aim to cater for these assumptions. Now, consider the new tables you're creating and their columns. Add SQL statements to your script that create indexes to facilitate fast retrieval on human-readable information they contain.
3. Test and debug your script until you're confident that it fulfils its purpose, i.e., to reflect the schema for the extended *Orinoco* database (as you developed it in Steps 1-2 in Part 1) and to populate this extended database with the instance-level information as described in Part 2.

Examples Clinic 4

Normalising Relational Schemas

Specification

The goal of this examples clinic is to provide you with some practice on normalising relational schemas. It aims to complement the lab exercise in which you are asked to map the EER model for the Orinoco database into a relational database schema.

DURATION: 50 minutes examples clinic time.

DESCRIPTION: The examples clinic consists of two tasks: in the first one you will examine a relation and, given the functional dependencies that hold within it, you need to decide which normal form the relation is in, explaining your answer, and then, in the second you need to normalise the relation, justifying each decomposition step. Work through the questions, demonstrate them to the lecturers and TAs in order to obtain feedback.

TASKS

T1. Consider the following relation about published books:

BOOK (Book_title, Authorname, Book_type, Listprice, Author_affil, Publisher)

Note that attribute Author_affil refers to the affiliation of the author (e.g., the university she or he belongs to).

Assume that the following dependencies hold in Book:

Book_title \rightarrow {Publisher, Book_type}
 Book_type \rightarrow Listprice
 Author_name \rightarrow Author_affil

1. State what normal form the relation is in and explain your answer.
2. Apply normalisation until you cannot decompose the relations further, justifying each decomposition step.

T2. Consider the collection of relations and dependencies (see (1)-(5) below).

Assume that each relation is obtained through decomposition from a relation R with attributes ABCDEFGHI and that all the known dependencies over relation R are listed for each item below. (The items are independent of each other, obviously, since the given dependencies over R are different.)

For each derived relation, state the strongest normal form that the relation is in and, if it is not in BCNF, decompose it into a collection of BCNF relations.

1. R1(A,C,B,D,E) $A \rightarrow B, C \rightarrow D$
2. R2(A,B,F) $AC \rightarrow E, B \rightarrow F$
3. R3(A,D) $D \rightarrow G, G \rightarrow H$
4. R4(D,C,H,G) $A \rightarrow I, I \rightarrow A$
5. R5(A,I,C,E)

Acknowledgements

Task T2 in this practical session is a minor adaptation of material authored and made available to instructors by Raghu Ramakrishnan and Johannes Gehrke to accompany their textbook Database Management Systems, 3rd Edition, McGraw-Hill, 2002, 978-0-072-246563-1 Copyright © 2002 McGraw-Hill, whom we thank. All errors are our responsibility.

Lab Exercise 04

Making Use of a Relational Database Specification

Overview

The goal of this lab exercise is to provide you with an opportunity to gain some practice in writing and executing more complex queries (e.g., queries containing joins, aggregates, such as COUNT, and Views.).

You will start by building language-neutral queries using relational algebra (RA) and mapping your RA expressions into simple SQL Queries. Then, you will be requested to write more complex SQL queries and decompose scenarios, whose descriptions are provided to you, into SQL queries. To complete this lab exercise, you will need to understand the relational model, the syntax of SQL (including the syntax of update and insert), and index-creation concepts.

ATTENDANCE:	Compulsory
MARKS AVAILABLE:	20 marks.
DEMONSTRATION:	You are required to submit two files: a report in PDF format generated from the template we provide and an archived folder of the outcomes of your work (scripts, spooling output, and diagrams).
TIME AVAILABLE:	2 hour lab session.
DEADLINE:	The end of the session.
EXTENSIONS:	Provided you attended the lab session and worked on the exercise all the time, you are granted an automatic extension. No further extensions are granted without documented special circumstances. This is strictly enforced.

Introduction

This lab exercise is composed of three parts: in **Part 1**, you are requested to create language independent relational algebra expressions for a number of simple selection queries; in **Part 2**, you are requested to write and execute more complex SQL expressions, containing joins, aggregates and other modifiers; in **Part 3**, you are requested to write and execute SQL queries based on the description of scenarios provided to you.

Requirements

From a Linux machine, you will use SQL*Plus to access the Oracle DBMS. The instructions for that were given in Lab Exercise 1. You should use the same user-name and password as you used in Lab Exercise 1.

Instructions

Firstly, work your way through Part 1. This should result in a set of RA expressions for all the queries in that part. Then, work your way through Part 2. This time no RA expressions result, only SQL queries that, once more, you need to ensure you obtain the desired results when you evaluate them against the database. Finally, work your way through Part 3, for which the outcome is similar to Part 2, i.e., the SQL statements that capture the requirements and the results that those statements produce when you evaluate them against the database.

Submission Contents

Your submission must contain:

1. The RA expressions required in Part 1 along with a plain-English description of each expression, stating what you think your expression says [7 marks]
2. The SQL queries requested in Part 2, with the evaluation results [7 marks]
3. The SQL queries requested in Part 3, with the evaluation results [4 marks]

Additional marks are available as follows:

4. Good report presentation (see the **Lab Sessions/Example Clinics** section above) [2 marks]

Part 1: Writing simple queries as RA expressions and translating them into SQL

Starting from your extension to the *Orinoco* database (as resulted from your work on Lab Exercise 3), you need to write down language-neutral RA expressions for each of the queries below. As an **optional task**⁶ try to translate the RA expressions into SQL queries over the schema that resulted from your work on Lab Exercise 3. You should run each query, check the results it produces and then, if necessary, debug either (or both) the queries and the schema as appropriate.

Remember that it is best to use a text editor to write then either copy and paste it into SQL*Plus or invoke the .sql file as a script. Also, remember that it is often possible, and desirable, to build a query by stages, incrementally.

The queries are:

- I. Select all columns/fields from each of the following tables:
 - A. manager
 - B. manager phone
 - C. artist
 - D. solo artist (as you named it in Lab Exercise 3)
 - E. group artist (as you named it in Lab Exercise 3)
 - F. sound engineer
- II. Select all columns/ fields from the following tables:
 - A. solo artist (as you named it in Lab Exercise 3), just for entities with the artistic name of Palmer John;
 - B. group artist (as you named it in Lab Exercise 3), just for entities called Goldfrat;
 - C. master track, just for entities edited by Max Spring.
 - D. hasMember, just for entities of the group Goldfrat.
- III. Project the names of solo artists that are members of the group Flut.
- IV. Select all columns/fields from master track, in two ways, as follows:
 - A. ordered by working title in ascending order
 - B. ordered by working title in descending order.
- V. Project the columns 'recordedBy' and 'working_title' from master track, for the tracks edited by Max Spring, ordered by working title in reverse order.

Part 2: Writing and executing complex SQL queries

Write SQL queries for each of the following retrieval requests and then execute them to check they compute the right answer.

1. Select all columns/fields from master track and sound engineer joining them on the sound engineer's ID and using an alias for each table.
2. Find all solo artists who are members of Goldfrat, only printing the columns in the solo artists table and using an alias for each table.
3. Display each phone number joined to its associated manager's name (use an outer join for this).
4. Count the number of tracks in the master track table.
5. List the artists that are members of either Goldfrat or Scandal (or both).
6. Who belongs to both Goldfrat and Scandal?
7. Who are the members of Goldfrat that do not belong to Scandal?
8. Create a view selecting all the members of Scandal.
9. Create a view selecting all the members of Goldfrat.
10. Compute the intersection of the views in (8) and (9) above.
11. Create a nested query to select the price of the catalogue entry with the highest amount of stock.
12. Create a nested query to select all buyers that have placed an order.

⁶Things to think about. There are no marks associated with this task.

13. Create a view to return the album information for the CD titled *20 People in a Field*.
14. Now use this view ((13) above) as part of a nested query to show the sequence in which the finished tracks appear as well as their release title.

Part 3: Using SQL queries to model data requirement scenarios

Write SQL statements that meet the requirements stipulated in the following scenarios:

1. Often a buyer will wish to see a list of tracks on an album in the order they appear. One such buyer wants to see the play list for the CD called *My Feet*. You were asked to write a SQL query that fulfils this requirement taking into account the constraint that your manager told you to only use joins to accomplish it.
2. If buyers wish to purchase a catalogue item, they are only allowed to do so if there is enough stock. If a purchase can be made, the stock in the catalogue must be updated to reflect that purchase. You should create a single SQL statement that performs both tasks, updating the catalogue only if the desired item exists and if there is enough stock. Use the CD *20 People in a Field* as your test case.

Examples Clinic 5

Advanced SQL and Transactions

Specification

The goal of this examples clinic is to provide you with some practice on writing triggers and procedures, and on reasoning about transactions. It aims to complement the lab exercise in which you are asked to write triggers and procedures over the Orinoco database.

DURATION: 50 minutes examples clinic time.

DESCRIPTION: The examples clinic consists of two tasks: in the first one you will practice writing a trigger and a procedure, and then, in the second you practice writing concurrent schedules given serial ones. Work through the questions, demonstrate them to the lecturers and TAs in order to obtain feedback.

TASKS

T1.

- I. What is PL/SQL?
- II. Write a procedure called `test1` to insert the variables `firstname`, `surname`, `age`, and `salary` into a table called `staff`.
- III. What is a trigger and what can triggers be used for?
- IV. Consider the tables shown below. The `numBookings` table stores the number of bookings (`num`) for a given train (`trainnum`) departing on a given date (`traindate`), where the number of bookings for a train on a date is obtained by counting the number of tuples in the `booking` table, which associates a given train (`trainnum`) and date (`traindate`) with a customer (`custnum`). Now, write a trigger that monitors inserts in the `booking` table and updates the `numBookings` table after each insert, keeping the number of bookings for each train up-to-date.

```
CREATE TABLE numBookings (
    trainnum    VARCHAR(10) references train(trainnum),
    traindate   DATE,
    num         NUMBER,
    primary key (trainnum, traindate)
);
```

```
CREATE TABLE booking (
    custnum     VARCHAR(10) references customer(custnum)
                on delete cascade,
    trainnum    VARCHAR(10) references train(train#),
    traindate   DATE,
    primary key (custnum, trainnum, traindate)
);
```

T2.

- I. Consider transactions T_1 and T_2 described below, where $r_n(x)$ indicates a read operation on data item x by transaction T_n , and $w_n(x)$ indicates a write operation on data item x by transaction T_n :

Transaction T_1 : $r_1(x)$ $w_1(y)$
 Transaction T_2 : $w_2(z)$

Also, consider the following serial schedules for transactions T_1 and T_2 :

Serial schedule T_1, T_2 : $r_1(x)$ $w_1(y)$ $w_2(z)$
 Serial schedule T_2, T_1 : $w_2(z)$ $r_1(x)$ $w_1(y)$

Create a concurrent schedule that is equivalent to both serial schedules.

- II. Similarly, consider transactions T_1 and T_2 described below, where $r_n(x)$ indicates a read operation on data item x by transaction T_n , and $w_n(x)$ indicates a write operation on data item x by transaction T_n :

Transaction T_1 : $r_1(z)$ $r_1(q)$ $w_1(y)$
 Transaction T_2 : $r_2(q)$ $w_2(z)$

Also, consider the following serial schedules for transactions T_1 and T_2 :

Serial schedule T_1, T_2 : $r_1(z)$ $r_1(q)$ $w_1(y)$ $r_2(q)$ $w_2(z)$
 Serial schedule T_2, T_1 : $r_2(q)$ $w_2(z)$ $r_1(z)$ $r_1(q)$ $w_1(y)$

Create a concurrent schedule that is equivalent to the serial schedule T_1, T_2 .

Is the schedule you created also equivalent to the serial schedule T_2, T_1 ? Why or why not?

Lab Exercise 05

Advanced SQL: Triggers and Procedures Specification

Overview

The goal of this lab exercise is to provide you with an opportunity to gain some practice in writing and executing triggers and PL/SQL procedures. To be able to solve this exercise, you will need to understand what triggers are and how they can be used to enforce integrity; in addition, you will need to understand how to create PL/SQL programs by using a combination of SQL and control statements, including the use of cursors.

ATTENDANCE:	Compulsory
MARKS AVAILABLE:	20 marks.
DEMONSTRATION:	You are required to submit two files: a report in PDF format generated from the template we provide and an archived folder of the outcomes of your work (scripts, spooling output, and diagrams).
TIME AVAILABLE:	2 hour lab session.
DEADLINE:	The end of the session.
EXTENSIONS:	Provided you attended the lab session and worked on the exercise all the time, you are granted an automatic extension. No further extensions are granted without documented special circumstances. This is strictly enforced.

Introduction

This lab exercise is composed of two parts: in **Part 1**, you are requested to write and test two simple triggers to support the maintenance of a derived attribute in the *Orinoco* database; in **Part 2**, you are requested to write and execute two PL/SQL procedures to support the maintenance of database integrity constraints and one in support of generate a formatted listing.

Requirements

From a Linux machine, you will use SQL*Plus to access the Oracle DBMS. The instructions for that were given in Lab Exercise 1. You should use the same user-name and password as you used in Lab Exercise 1.

In addition to the lecture handouts and the mandatory readings, it could be useful for you to consult reputable, trustworthy online sources for information on PL/SQL. Here is one recommendation that is available online to you as a University of Manchester student:

Oracle and PL/SQL Recipes
A Problem-Solution Approach
Josh Juneau, Matt Arena
ISBN: 978-1-4302-3207-0 (Print) 978-1-4302-3208-7 (Online)
<http://link.springer.com/book/10.1007/978-1-4302-3208-7>

Note, as always, that, often, there will need to be some adjustments due to the local setting of our Oracle DBMS.

Instructions

Work your way through each of the three parts above.

Submission Contents

Your submission must contain:

1. The code for the trigger in Part 1 with the evaluation results⁷ [4 marks]
2. The code for the procedures and view in Part 2 with evaluation results:
 - 2.a The PL/SQL procedure for the ContractInfo [4 marks]
 - 2.b.i. The AlbumDistribution view [4 marks]
 - 2.b.ii. The PL/SQL procedure over the view from 2.b.i [6 marks]

Additional marks are available as follows:

3. Good report presentation (see the **Lab Sessions/Example Clinics** section above) [2 marks]

Part 1: Writing triggers

In this lab exercise (for all parts), remember that it is best to use a text editor to write then either copy and paste it into SQL*Plus or invoke the .sql file as a script. Also, remember that it is often possible, and desirable, to build a query by stages, incrementally. One additional thing to pay attention to here is that when you're debugging your triggers (and procedures) later on, an error may leave the database in an unexpected and undesirable state. Because of this, always think it through that you need to take action to reverse the changes and how to do so. Finally, consider carefully the need to add a single slash in as the first and single character in the final line of a trigger or procedure.

Starting from your extension to the *Orinoco* database (as resulted from your work on Lab Exercise 3), you need to write a single trigger that fulfils the following requirements:

- The duration column in the ContractInfo table stores a derived attribute (you can see this on the EER diagram you were given). Create a trigger to compute the value of the duration column when a tuple is inserted in ContractInfo. Also, make sure that the trigger computes the value of the duration column when a tuple in ContractInfo is updated.

In doing this task, it is best to:

- retrieve all the tuples in the table (so that you know the state before the trigger is fired), then
- perform an operation that fires the trigger (e.g., `INSERT INTO ContractInfo VALUES ('Goldfrat', '10-oct-2008', '09-oct-2010', 0)`; for the first part of the trigger above and `UPDATE ContractInfo SET duration=0`; for the second part of the trigger above, then, finally,
- retrieve again all the tuples in the table (so that you can see and compare the state after the trigger has fired with that that existed before).

In your report, you must necessarily include the trigger and its results. As always, also include the output of the SQL script you created as part of the PDF document. In addition, make sure that you include the SQL script(s), within the .zip file you are uploading in Moodle, that causes the creation of the trigger, and make sure when it runs it produces the results from your report.

Part 2: Writing procedures

Starting from your extension to the *Orinoco* database (as resulted from your work on Lab Exercise 3), you need to write SQL statements and PL/SQL procedures, as follows:

- a) Create a single PL/SQL procedure that wraps the insertion of tuples in the ContractInfo table so as to make sure that the start and end dates of the new tuple do not lie between the start and end dates of any contract already present in the table, and that the from and to dates are in the correct temporal order.

⁷ Here and elsewhere, in your report, aim to show in your report the log of Oracle runs, e.g., as obtained by spooling, which we discussed in Lab Exercise 1. For this lab exercise, only a single correct run for each trigger or procedure is needed to help the marking.

b) Write SQL statements and PL/SQL procedures that meet the requirements:

- (i) Create a view, named it `AlbumDistribution`, to return the album information along with a new attribute called `is_distributed_as` to store information as to whether an album can be distributed as a CD, on vinyl, or as a cassette tape (e.g., 'c', 'v', or 't').

- (ii) Recall the track listing scenario from Lab Exercise 4. It stated:

Often a buyer will wish to see a list of tracks on an album in the order of the play list. One such buyer wants to see the play list for a CD called "My Feet". It is your job to fulfil this request but your manager only wants you to use joins to accomplish it.

Now, use the view (2.b.i from above) and create a PL/SQL procedure that extends and wraps the resulting as follows. Firstly, it must take as input parameters the album type (e.g., 'c', 'v', or 't') and the album title. Secondly, in the case of album title, it must allow one to enter wildcard sequences, (e.g., '%e%', or '%t%'). Finally, it must produce a formatted listing, with the sequence followed by a comma, then the title of the track followed by a greater-than (i.e., right angle-bracket), followed by the album title. **Hint:** remember that you need to set the server output on.

In doing these tasks, it is best to:

- retrieve all the tuples in the table (so that you know the state before the procedure is invoked for execution), then
- invoke the procedure with values that test the logic of your code (e.g., in the the case of the first procedure, use one or two INSERTs that are illegal and one or two that are legal, and so on), then, finally,
- retrieve again all the tuples in the table (so that you can see and compare the state after the procedure has executed with that that existed before).

In your report, you must necessarily include the PL/SQL procedures as well as the SQL statements that create the view. As always, include the output of the SQL script you created as part of the PDF document. In addition, make sure that you include the SQL script(s), within the .zip file you are uploading in Moodle that causes the creation of the trigger, and make sure when it runs it produces the results from your report. Do study the `ReadMe.txt` file for detailed instructions.