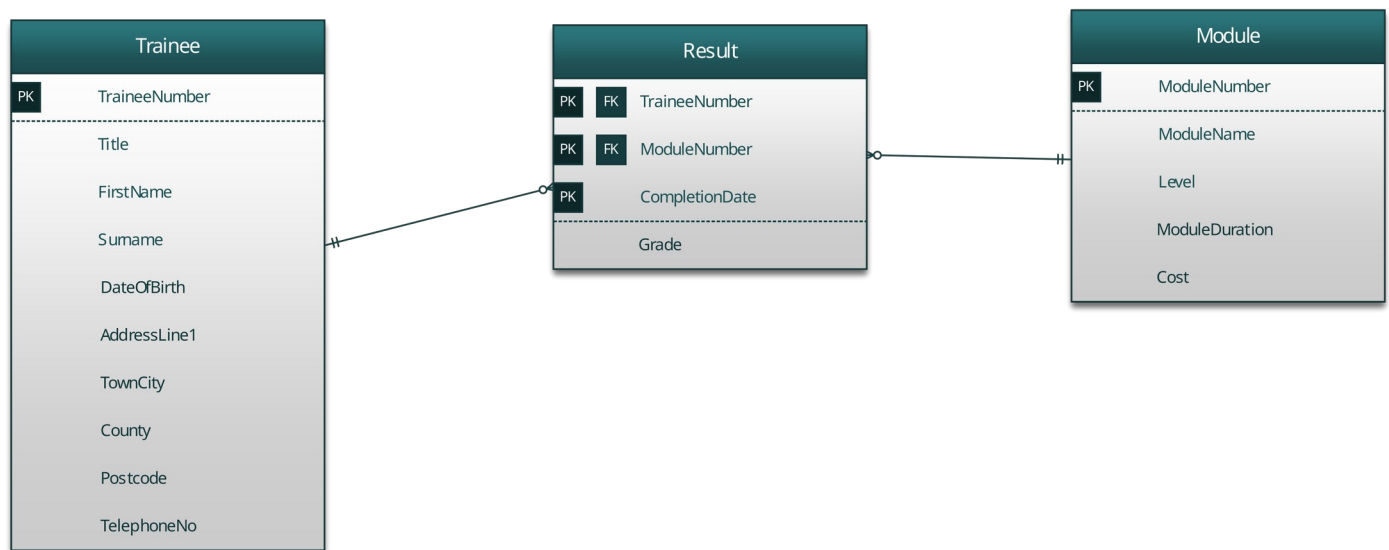


S2.6: DATA INTEGRITY

TASK 1: Creating the Data Dictionaries

We are looking at a database for a Circus School, that stores information about Trainees, Modules that can be studied, e.g. fire eating, juggling, etc., and Results for the trainees taking each module. The data has been normalised to 3NF, and is stored in three tables: Modules, Results and Trainees as below. You can see the actual data in the Circus Data spreadsheet provided on Blackboard.



Complete the data dictionaries below for each table. Validation should ensure that the fields are all of the correct type and size, which will help to minimise erroneous data and maximise use of disk-space. Create the validation and verification checks you think would help stop data errors being entered. You do not need to know how to implement all of it – just how you think the data could be validated. Please check the data in the spreadsheet to see the actual data, which will give you an idea of the validation and verification that would be possible.

Notes on the column headings:

Data type: text, date, number, currency, Boolean, etc.

Field size: Text = the number of characters, Date/time = short, med, long, Number = integer, double, decimal, etc.

Other validation: input masks (postcodes, email addresses, etc.), primary & foreign keys, formatting, look-ups, pick-lists, etc.

Modules Table

(examples have been given to start you off, but you should expand if necessary)

Field Name	Data Type	Field Size	Other Validation/Information
ModuleNo	Char	3	Numbers only
ModuleName	Varchar	30	
Level	Char	1	Only between 1 and 3. First number of Module No
Duration	Char	2	Only numbers allowed
Cost	Number	Integer	>=0

Results Table

Field Name	Data Type	Field Size	Other Validation/Information
TraineeNumber			
ModuleNo			
CompletionDate			
Result			

Trainee Table

Field Name	Data Type	Field Size	Other Validation/Information
TraineeNumber			
Title			
FirstName			
Surname			
DateofBirth			
AddressLine1			
TownCity			
County			
Postcode			
TelephoneNo			

I would suggest that you use the following as primary keys for the tables. Remember the primary key is necessary to ensure that each row in the table is unique:

Modules: **ModuleNo** (so that each module is recorded once)

Trainee: **TraineeNumber** (so that each trainee is recorded once)

Results: **TraineeNumber, ModuleNo and CompletionDate** (so that each trainee can be assessed for any module only once per day (they could fail it one day and retake another day))

TASK 2: Creating the Tables and Importing the Data

1. You will find the basic Create Table scripts on Blackboard to create each of the three tables. Based on your Data Dictionaries, and the validation/verification you have stated, edit the scripts for each of the tables and import them to your Oracle Apex workspace.

- a. Use some of the validation/verification on your tables – I am not expecting you to do all of it, as some of it can be quite complicated, e.g. such as a postcode format. Just do what you can.
 - b. Double check your validation against the Circus spreadsheet. In the next step we will import the data, and if any of the data violates your validation, you will get an error message during the import.
 - c. Remember that the order you run the scripts is important. You won't be able to create the Result table until you have created the Trainee and Module tables. This is because the Result table has foreign keys which reference the other two tables.
2. Now you have the three tables created in Apex, you are going to upload the data provided in the Circus spreadsheet to your three tables. Again, the order is important – you will need to upload the data to the Trainee and Module tables before the Result table:
- a. Firstly, download the Circus Data Excel file from Blackboard.
 - b. In Oracle Apex, navigate to SQL Workshop – Utilities – Data workshop, - click the 'Load Data' button, and then select the Circus Data file.
 - c. On the Load Data Form change 'Load To' from 'New Table' to 'Existing table'.
 - d. Then select the Module table from the drop down under 'Table', and in the 'Settings' area under 'Select Sheet' select Module. By doing this, we are telling Apex that we want to import the data on the Module sheet in the spreadsheet into the Module table in our database.
 - e. You now need to map the columns across from the spreadsheet (Module) to the columns in the Oracle table. Click the 'Configure' Button. On the next form you will see a list of the columns in the spreadsheet – 'Source Column', along with its Data Type. On the 'Map To' section Oracle may have attempted to map the columns to the table columns for you. But check they are the correct column, and any that haven't been mapped select the correct column from the list like this:

Source Column	Data Type	Map To	Format
MODULE_NUMBER	NUMBER	MODULENO (Char)	
MODULE_NAME	VARCHAR2(100)	MODULENAME (Varchar2)	
LEVEL_	NUMBER	MODULELEVEL (Char)	
MODULE_DURATION_DAYS_	NUMBER	MODULEEDURATION (Char)	
COST	NUMBER	MODULECOST (Number)	

- f. When you have mapped the columns click the 'Save Changes' button. And then on the next page click the 'Load Data' button. Hopefully you will get a green tick and a message saying that the Module table has had 11 new rows added. Click the 'View Data' button to see your table.

- g. If you have an error, it is because you have made a mistake with your verification/validation, and you are trying to import data that is not allowed, e.g. trying to add a telephone number with 13 characters where you have constrained the column to only allow 11 characters. If this is the case, you will need to double check your create table scripts against the data in the spreadsheet and adjust the scripts accordingly.
 - h. Repeat the above process for the Trainee table and then for the Result table, making sure you select the correct sheets in the spreadsheet and the correct tables in the database.
3. Finally, you should add some new data manually:
- a. Add 2 new trainee records.
 - b. Add 1 new module of your choice.
 - c. Finally add 5 training records for trainees who have taken the new module.

TASK 3: Querying the Data

1. List all trainees who live in Preston.
2. List all trainees that are under 30 years old (approximately is fine, so you can base it on 365 days per year).
3. List the details about all of the level 3 modules.
4. List the module name and duration of modules that cost more than £50.
5. List all trainees who have failed a module in the last 2 months.
6. List the module that has the most trainees passing with a distinction.
7. Which module has the longest name.
8. If there are module(s) that have never been taken by any trainee, list them.
9. If there are any trainees that have never taken a module, list them.
10. List the details of the trainees who haven't got a county entered on their record.
11. List the first name, surname and date of birth of all trainees in order of their age.
12. List the trainee number, first name and surname of all trainees who have taken a level 2 module.
13. List the trainee first name, surname, the module title, the date of completion and their result.
14. How much has Rev. Brian Broadgate spent on Circus modules in total?
15. Calculate the total spent on modules for each trainee. List in descending order of who spent the most money.
16. Calculate the total money made from each module. List in descending order of the one that has brought in the most money.

Note that any work not finished in class should be completed in your own time.

Also, if you are behind with any work from previous weeks, please spend any remaining time finishing it off. This will be very beneficial to you when we come to exam time!