LEVEL 5

DATABASE DESIGN AND DEVELOPMENT

Student Guide

Modification History

Version	Date	Revision Description
V1.0	October 2011	For release
V1.1	November 2015	Assessment Methodology Updated

Published by: NCC Education Limited, The Towers, Towers Business Park, Wilmslow Road, Didsbury, Manchester M20 2EZ, UK.

Tel: +44 (0) 161 438 6200 Fax: +44 (0) 161 438 6240 Email: info@nccedu.com
http://www.nccedu.com

CONTENTS

1. Module Overview and Objectives

This module builds on a basic understanding of database concepts to develop your skills in the design and development of databases and database management systems, as well as investigating enterprise applications of databases.

2. Learning Outcomes and Assessment Criteria

Learning Outcomes; The Learner will:	**Assessment Criteria;** The Learner can:
1. Understand the enterprise application of database systems	1.1 Summarise the common use of distributed database management systems 1.2 Explain the meaning of the term disributed database management system 1.3 Describe the components of a disributed database management system 1.4 Summarise the common use of data warehouses 1.5 Explain the meaning of the term data warehouse 1.6 Describe the structure of a data warehouse
2. Understand how to enhance the design of and further develop a database system	2.1 Describe how tables that contain redundant data can suffer from update anomalies 2.2 Explain how to overcome update anomalies using normalisation 2.3 Describe how to retrieve data from one or more tables using SQL
3. Be able to enhance a logical database design	3.1 Check the tables are well-structured using normalisation 3.2 Define the integrity constraints on the tables
4. Be able to develop a physical database design	4.1 Map a logical database design to a physical database design 4.2 Design tables for a target DBMS 4.3 Design a representation of derived data 4.4 Design integrity constraints for the target DBMS 4.5 Denormalise tables where appropriate
5. Be able to enhance a database system using SQL	5.1 Apply integrity constraints 5.2 Retrieve data from one or more tables using join 5.3 Retrieve data from one or more tables using sub-queries

3. Syllabus

Syllabus			
Topic No	Title	Proportion	Content
1	Key Concepts in Databases and Database Development	1/12 2 hours of lectures 2 hours of tutorials 1 hour of laboratory sessions	• Review of key material from Level 4 databases module • Common uses of databases • Types of databases • Overview of database development ***Learning Outcome: All***
2	Enhancing Design 1	1/12 2 hours of lectures 2 hours of tutorials 1 hour of laboratory sessions	• Introduction to normalisation • The concept of functional dependency • Data redundancy and update anomalies • Overcoming anomalies with normalisation ***Learning Outcome: 2***
3	Enhancing Design 2	1/12 2 hours of lectures 2 hours of tutorials 1 hour of laboratory sessions	• Deriving a set of relations from a conceptual data model • Validating relations using normalisation • Integrity constraints on tables ***Learning Outcome: 3***
4	Data Retrieval 1	1/12 2 hours of lectures 2 hours of tutorials 1 hour of laboratory sessions	• Table and view structure in a relational database • Data types • Null values • Retrieving data using SQL ***Learning Outcome: 2***

5	Data Retrieval 2	1/12 2 hours of lectures 2 hours of tutorials 1 hour of laboratory sessions	• Referential integrity in relational databases • Types of joins • Retrieving data using joins • Retrieving data using sub-queries ***Learning outcome: 5***
6	Physical Design 1	1/12 2 hours of lectures 2 hours of tutorials 1 hour of laboratory sessions	• The purpose of physical design • Mapping the logical database design to a physical database design • Designing tables for the target DBMS ***Learning Outcome: 4***
7	Physical Design 2	1/12 2 hours of lectures 2 hours of tutorials 1 hour of laboratory sessions	• The concept of derived data • Designing a representation of derived data ***Learning Outcome: 4***
8	Physical Design 3	1/12 2 hours of lectures 2 hours of tutorials 1 hour of laboratory sessions	• Types of constraints • Designing integrity constraints for the target DBMS ***Learning Outcome: 3 & 5***
9	Physical Design 4	1/12 2 hours of lectures 2 hours of tutorials 1 hour of laboratory sessions	• Understanding transactions • Denormalisation • Improving performance • Estimating the size of the database ***Learning Outcome: 4***

10	Distributed Databases	1/12 2 hours of lectures 2 hours of tutorials 1 hour of laboratory sessions	• The need for distributed databases • Components of distributed databases • Advantages and disadvantages of distributed databases • Homogenous and Heterogeneous distribution • Distributed Database Design ***Learning Outcome: 1***
11	Data Warehouses	1/12 2 hours of lectures 2 hours of tutorials 1 hour of laboratory sessions	• The need for business intelligence and the concept of the data warehouse • The difference between Online Transaction Processing (OLTP) systems and data warehousing • The architecture and main components of a data warehouse ***Learning Outcome: 1***
12	Summary	1/12 2 hours of lectures 2 hours of tutorials 1 hour of laboratory sessions	• Summary of module, linking units to objectives and to each other • Clarification of material and related issues as identified by students ***Learning Outcome: All***

4. Related National Occupational Standards

The UK National Occupational Standards describe the skills that professionals are expected to demonstrate in their jobs in order to carry them out effectively. They are developed by employers and this information can be helpful in explaining the practical skills that you have covered in this module.

Related National Occupational Standards (NOS)
Sector Subject Area: 6.1 ICT Professionals **Related NOS:** 4.1.P.1 – Contribute, under supervision, to the preparation of a data analysis assignment; 4.1.P.2 – Assist in the development of data analysis models; 4.1.P.3 – Manage the outcomes from the data analysis assignment; 4.5.P.2 – Manage, under supervision, the maintenance of data design assignments; 4.5.P.1 – Provide others, when requested, with specified information relating to data design activities; 4.5.S.1 – Select and implement appropriate data design processes; 4.5.S.2 – Manage the progress of data design assignments; 4.5.S.3 – Review the effectiveness of data design deliverables.

5. Teaching and Learning

Suggested Learning Hours					
Lectures:	**Tutorial:**	**Seminar:**	**Laboratory:**	**Private Study:**	**Total:**
24	24	-	12	90	150

The teacher-led time for this module is comprised of lectures, laboratory sessions and tutorials. You will need to bring this Student Guide to all classes for this module. The breakdown of the hours is also given at the start of each topic.

5.1 Lectures

Your lecturer will be presenting the basic knowledge and the theoretical concepts required for the unit during this time. He/she will use PowerPoint slides during the lecture time and you will be expected to take notes.

You will also be encouraged to be active during this time and discuss and/or practice the concepts covered. Lectures will include question and answer elements to promote participation and to allow your lecturer to check whether you understand the concepts they are covering.

5.2 Tutorials

These are designed to deal with the questions arising from the lectures and private study sessions. You should think carefully beforehand about any areas in which you might need additional guidance and support and use this time to discuss these with your teacher.

5.3 Laboratory Sessions

During these sessions, you are required to work through practical tutorials and various exercises. The details of these are provided in this guide.

5.4 Private Study

This Student Guide also contains details of the private study exercises. You are expected to complete these exercises to improve your understanding. Your tutor will set deadlines for the completion of this work and go over the suggested answers with you. The deadlines will usually be before the scheduled tutorials for that topic. Some of the private study tasks may require you to work in a small group so you will need to plan your time carefully and ensure that you can meet with your group members to complete the work required before the deadline.

You should also use this time to revise the content of lectures to ensure understanding and conduct extra reading (using the supplementary textbooks or other materials available in the library or online). You should bring any questions to the tutorial for additional guidance and support.

6. Assessment

This module will be assessed by means of an assignment worth 50% of the total mark and an examination worth 50% of the total mark. These assessments will be based on the assessment criteria given above and you will be expected to demonstrate that you have met the module's learning outcomes.

Assignments for this module will include topics covered up to and including Topic 7. Questions for the examination will be drawn from the complete syllabus.

7. Further Reading List

You will also be expected to undertake further reading to consolidate and extend your knowledge of the topics covered in this module. Your Accredited Partner Centre's library will contain a selection of useful sources of information and you can also make use of materials available online. The list below also provides suggestions of suitable reference books you may like to use:

Benyon-Davies, P. (2003). *Database Systems,* 3rd edition*.* Palgrave Macmillan.
ISBN-10: 1403916012
ISBN-13: 978-1403916013

Connolly, T. and Begg, C. (2003). *Database Solutions: A step-by-step guide to building database*, 2nd edition. Pearson Addison-Wesley.
ISBN-10: 0321173503
ISBN-13: 978-0321173508

Connolly, T. and Begg, C. (2014). *Database Systems: A Practical Approach to Design, Implementation and Management,* 6th edition*.* Pearson Addison Wesley.
ISBN-10: 0132943263
ISBN-13: 978-0132943260

Dietrich, S. (2001). *Understanding Relational Database Query Languages.* Pearson Prentice Hall.
ISBN-10: 0130286524
ISBN-13: 978-0130286529

Kroenke, D. (2013). *Database Processing: Fundamentals, Design and Implementation,* 13th edition*.* Prentice Hall.
ISBN-10: 0133058352
ISBN-13: 978-0133058352
.

Topic 1: Key Concepts in Databases and Database Development

1.1 Learning Objectives

This topic provides an overview of the module as a whole, a review of key material from Level 4 and an introduction to the stages of database design and development.

On completion of the topic, you will be able to:

- Understand which topics will be covered in the module;
- Recognise key material from the Level 4 Databases module;
- Outline common uses of databases and different types of databases;
- Give an overview of database development.

1.2 Timings

Lectures:	2 hours
Laboratory Sessions:	1 hour
Tutorials:	2 hours
Private Study:	7.5 hours

1.3 Laboratory Sessions

The laboratory time allocation for this topic is 1 hour.

These laboratory sessions will involve you in creating two sets of tables. In this session, you will create the students and courses tables that will be used throughout the lab sessions in later topics.

Later in the module you will create some tables around a scenario called ‘Marlowe Interiors’. These will also be used in some laboratory sessions.

Creating Tables

You should remember how to create tables and insert data from previous study Here is an example from the Level 4 Databases module exercises:

```
Create table departments
(dept_no  integer not null primary key,
department_name varchar(30),
location varchar(20);

Create table workers
(emp_no integer not null primary key,
first_name varchar(30),
last_name varchar(30),
job_title varchar(30),
age integer,
dept_no integer,
foreign key (dept_no) references departments)
```

Exercise 1:

Create two tables, 'Courses' and 'Students'.

The 'Courses' table should have a course_id which is an integer, and a course_name which is a varchar 30 long. Course_id will be the primary key.

The 'Students' table should have a student_id, an integer which is the primary key.

Other attributes:

```
first_name varchar(30)
last_name varchar(30)
student_type varchar(20)
```

course_id is a foreign key that references course_id on the courses table.

Exercise 2:

Inserting data into the two tables

The format of an insert statement is shown in this example from the Level 4 module.

```
Insert into departments values ('1','Packing','Cairo');
```

Using the same format use the Insert command to create the following courses:

1. Computing Science
2. History
3. Geography

Insert the following students using the SQL insert statements:

Remember that the course_code must reference one of the course codes you have just created in the courses table.

1. Pavel Dobovitch, Home Student, 1
2. Winston Kodogo, Overseas Student, 1
3. Dawn Cove, Overseas Student, 1
4. Satpal Singh, Home Student, 2
5. Horace Smith, Home Student, 3

Don't forget to commit your data!

1.4 Private Study Exercises

You should spend approximately 7.5 hours on the Private Study for this topic. You should use this time to complete the exercises below as directed by your lecturer and to review the contents of this topic.

Exercise 1:

Write a report in two parts.

The first part should give an overview of the three main phases of database development: conceptual design, logical design and physical design. You should outline the main activities of each of these phases. Research these phases using textbooks and web sources. Here are some examples of textbooks you could use, but you may also find your own resources. Please note, these are not required textbooks.

- Connolly, T. and Begg, C. (2005). *Database Systems A Practical Approach to Design, Implementation and Management*, 4th edition. Addison and Wesley. Chapters 1 and 9.
- Benyon-Davies, P. (2000). *Database Systems,* 3rd edition. Palgrave. Chapter 14.
- Kroenke, D. M. (2005). *Database Processing: Fundamentals, Design and Implementation,* 10th edition. Prentice Hall.
- Applied Information Science. (1997). *Conceptual, logical and physical models.* [Available Online] http://www.aisintl.com/case/CDM-PDM.html
- Nicewarner, N. (2004). Why we have conceptual, logical and physical data modelling? *Power Builders Developer's Journal,* 1 Feb 2004. [Available Online] http://pbdj.sys-con.com/node/106944

The second part should be on one of the additional aspects of database development. Your lecturer will allocate the topics to ensure that different content is covered evenly.

- DBMS selection
- Application design
- Transaction design
- Data Conversion
- Testing

Your report should be 600-900 words in length and you should be prepared to discuss your report in the tutorial session.

Exercise 2:

Read the following scenario about a company called Marlowe Interiors. This scenario will be used as an example in future private study, tutorial sessions and lab sessions.

> *Background on the firm*
>
> Marlowe Interiors is a trendy interior design and building firm located in Islington, North London. They specialise in 'makeovers' to individual rooms or entire properties. They evolved from the traditional building firm of Ted and Arthur Barrett, two brothers who had done all sorts of work in the area for many years. With the craze for makeovers taking off,

the brothers decided to seize the time and change the nature of the business. This involved recruiting some new workers, notably Horace the interior designer, and changing their image with a new name and logo. However old habits die hard and the accounts and records of the company were kept in the same old way, in files in Ted and Arthur's office. For example there is a card box with index cards for customer records and a box file with copies of all the old invoices that have been sent out. Now the brothers have decided that since the firm is doing very well then it is time to have a computerised record system that can hold all the information they need in one place.

Determine what entities and attributes you can identify. Also suggest what further materials or information from the firm would be useful in identify entities and attributes.

1.5 Tutorial Notes

The tutorials for this topic will last for 2 hours. You can expect to spend some of this time discussing your answers to the Private Study exercises with your lecturer and other students. Your lecturer will direct you on completing the tasks below.

Exercise 1:

In small groups based on the topic that you have been given, discuss the topic and develop a presentation for the rest of the class. The presentation should last approximately 5 minutes. There should be presentations on each of the following areas:

- DBMS selection
- Application design
- Transaction design
- Data Conversion
- Testing

Exercise 2:

Again in different small groups, discuss the following stages of database design and present the information to the rest of the class:

- Phases of database design: conceptual design
- Phases of database design: logical design
- Phases of database design: physical design

Topic 2: Enhancing Design 1

2.1 Learning Objectives

This topic provides an overview of normalisation. On completion of the topic, you will be able to:

- Explain the reason why the process of normalisation is carried out;
- Normalise a document to third normal form.

2.2 Timings

Lectures:	2 hours
Laboratory Sessions:	1 hour
Tutorials:	2 hours
Private Study:	7.5 hours

2.3 Laboratory Sessions

The laboratory time allocation for this topic is 1 hour. This session will use the students and courses tables created in the laboratory session for Topic 1.

Exercise 1:

Add a column to the students table to record a student's age.

Exercise 2:

These exercises serve as a recap of some basic SQL selections based on the tables created in the previous section.

1. Select first name and last name of all the students
2. Select the names of all the courses ordered by the course name
3. Select all the courses and the students that are in them
4. Select the last name and course name of all the overseas students
5. Show the course data and names of all the students studying history

Exercise 3: Greater than; less than. Not equals to.

SQL can use the following symbols:

Greater than

< Less than

<> Not equal to

>= greater than Or equal to

<= Less than or equal to

Use these structures to perform the following queries;

1. Select all the data for students over the age of 21
2. Select the first name, last name and age for students who are twenty one or over
3. Select the first name, last name, course id and course name for students who are not studying Computing Science
4. Select all the details for the students under 30 who are studying computer science
5. Show the first names for students studying Computing Science who are not Home Students

2.4 Private Study Exercises

You should spend approximately 7.5 hours on the Private Study for this topic. You should use this time to complete the exercises below as directed by your lecturer and to review the contents of this topic.

Exercise 1:

Read all the relevant material for this topic. This should including the slides and slide notes along with the sections from at least one relevant textbook.

You should make sure you understand the concepts of functional dependence and definitions of each normal form up to 3rd Normal Form.

You should make sure that you read some material on the existence of higher normal forms.

Prepare a short presentation that could be either a PowerPoint presentation or discussion document that describes:

a. Boyce-Codd Normal Form (BCNF)
b. Fourth Normal Form (4NF)

There should be a definition of each of these normal forms and an example. It is acceptable to use an example from a published source so long as a proper reference should be given. Be prepared to present your findings in the tutorial session for this topic.

Exercise 2:

Carry out a normalisation to 3rd Normal form on the job sheet from Marlowe Interiors shown below.

Draw the ER diagram that results from this normalisation.

7. JOB SHEET

JOB NUMBER: 0023
CUSTOMER NAME: Carol Smiley
CUSTOMER ADDRESS: 1 Vickers Street, N1
JOB DESCRIPTION: Room redecoration

Employee No	Employee Name	Worker Type	Worker Type Hourly Rate	Hours worked on Job	Total for Job
009	Arthur Barrett	Qualified Builder	55	2	110
011	Sid Jones	Labourer	20	10	200
012	Jane Dean	Plumber	50	1	50

020	Horace D'Ville	Interior Designer	39	1	39
019	Steve Crabapple	Labourer	20	11	220

2.5 Tutorial Notes

The tutorials for this topic will last for 2 hours. You can expect to spend some of this time discussing your answers to the Private Study exercises with your lecturer and other students. Your lecturer will direct you on completing the tasks below.

Exercise 1:

In small groups, discuss your findings from Private Study Exercise 1 and develop a group presentation that gives a definition and example of each of the higher normal forms asked for in the exercise.

Each group should present their ideas and the information should be discussed as a whole class group.

Exercise 2:

Discuss the following questions in small groups and feed back to the class as a whole:

a. What is the purpose of normalisation?
b. Explain the concept of functional dependency
c. Give a definition of normalisation up to 3rd normal form

Exercise 3:

Normalise the following invoice from Marlowe Interiors shown below and draw the resulting ER model.

Marlowe Interiors

1 Newington Green Road
London
8. N1 TYY
020 7888 1234

Job: Kitchen Makeover
Customer: Ivan Jones, 2 Digby Mansions. Highbury Park
Area: London North
Date: 02/03/00

Parts

1 sink, tin @ 130.00 including VAT.
1 u-pipe @ 20.00 ditto
3 x assorted plumbing fittings @ 33.00 total including VAT
1 thermostat @ 100.00 including VAT

Labour

Plumber 3 hours £150
Labourer 3 hours £60
Electrician 1 hour £50 (to fit thermostat)

TOTAL PARTS	*£283.00*
TOTAL LABOUR	*£260.00*
TOTAL	*£543.00*

Topic 3: Enhancing Design 2

3.1 Learning Objectives

This topic provides an overview of logical design. On completion of the topic, you will be able to:

- List the steps in logical design;
- Give an account of the main activities that make up logical design;
- Check the integrity constraints on their data model.

3.2 Timings

Lectures:	2 hours
Laboratory Sessions:	1 hour
Tutorials:	2 hours
Private Study:	7.5 hours

3.3 Laboratory Sessions

The laboratory time allocation for this topic is 1 hour. These exercises use the students and courses tables created in the laboratory sessions for Topic 1.

Exercise 1: Creating a table from another table and further queries

1. Create a table called student_types by selecting the distinct values of the student _type column from the students table.

2. Add a column to the student_types table to record the fees that will be charged.

3. Make the student_type attribute of student_types the primary key. You should use the alter table command

4. Use the alter table command to make the student_type attribute on students a foreign key to the new student_types table

5. Update the student_types table to set the fees. Home students will pay 5000, overseas students will pay 10000

6. Show each student's first and last names and the fees that they are paying.

3.4 Private Study Exercises

You should spend approximately 7.5 hours on the Private Study for this topic. You should use this time to complete the exercises below as directed by your lecturer and to review the contents of this topic.

Exercise 1:

Read all the relevant material for this topic. This should including the slides and slide notes along with the sections from at least one relevant textbook.

Ensure you have read and understood each of the steps in logical design that have been identified in the lecture. Prepare a presentation/discussion that outlines each of the steps of logical design. This could be either a PowerPoint presentation or discussion document.

Investigate the optional step (1.6) Specialise/Generalise in extra detail. This step relates to concepts in the Enhanced Entity Diagramming approach. Investigate this and discuss how this might be applied to the development process.

Be prepared to present this during the tutorial session.

Exercise 2:

Refer back to the Marlowe Interiors case study that you have worked with in the private study exercises in previous topics.

Draw the full entity relationship model from previous partial models that have resulted from normalisation in Topic 2. Be prepared to discuss this in the tutorial session.

You should incorporate the additional materials shown below:

1st Document
Rates of pay for various types of worker

9. Type	10. Rate per Hour
Plumber	£50
Labourer	£20
Qualified Builder	£55
Interior Designer	£39
Electrician	£50

2nd Document
Example of a manual customer record kept on index files in the filing cabinet

Name: Jane MacDonald Address: 23 Essex Road, N1 6YY Telephone: 070 222 3333 or 0603 99933 (mobile) or 0208 99999 (work) E-mail: janem@videotec.co.uk also jane@mc.eagle

3.5 Tutorial Notes

The tutorials for this topic will last for 2 hours. You can expect to spend some of this time discussing your answers to the Private Study exercises with your lecturer and other students. Your lecturer will direct you on completing the tasks below.

Exercise 1:

In small groups, discuss your findings from Private Study Exercise 1 and develop a group presentation that gives an overview of the steps of logical design and in particular the 'specialise/generalise' step.

Each group should present their ideas and the information should be discussed as a whole class group.

Exercise 2:

Discuss Private Study Exercise 2 as a whole class group. How did you go about creating the entity relationship model? Did everyone come up with the same model?

Exercise 3:

Discuss the following as a whole-class group:

a. What is meant by an iterative approach to the development of databases?
b. Why is it important to involve users in the development of a database?
c. What do you understand by the concept of a candidate key?
d. What is meant by the concept of data redundancy?

Exercise 4:

In small groups go through each of the steps of logical design and suggest how these might be applied to the development of the system for the Marlowe Interiors.

Be prepared to discuss this as a whole-class group.

Topic 4: Data Retrieval 1

4.1 Learning Objectives

This topic provides an overview of the basic aspects of data retrieval. On completion of the topic, you will be able to:

- Implement more complex relationships;
- Describe how to retrieve data from one or more tables using SQL;
- Recognise and identify different data-types in SQL.

4.2 Timings

Lectures:	2 hours
Laboratory Sessions:	1 hour
Tutorials:	2 hours
Private Study:	7.5 hours

4.3 Laboratory Sessions

The laboratory time allocation for this topic is 1 hour. These exercises use the Students and Courses tables created in the laboratory session in Topic 1.

Exercise 1:

Many important queries in a database involve using the aggregation functions COUNT, MIN, MAX, SUM and AVG

- COUNT counts the number of times something occurs in a database table, the number of rows that meet a particular condition.
- MIN finds the minimum or lowest occurrence of an attribute in a database table.
- MAX finds the highest occurrence of an attribute in a database table.
- AVG finds the mean average of an attribute in a database table.
- SUM finds the totals of all the values of an attribute in a database table

It is also useful to be able to use the aggregation functions with the group by and having clauses.

For example to show the number of students on each course:

```
select course_name, count(student_id)
from courses, students
where students.course_id = courses.course_id
group by course_name
```

We could modify this to shown only those courses with more than 1 student by using the having clause:

```
select course_name, count(student_id)
from courses, students
where students.course_id = courses.course_id
group by course_name
having count(student_id) > 1
```

Complete the following:

1. Show a count of students on each course and group by the course id
2. Using a sub-query show the first name, last name and age of the oldest student
3. Using a sub-query show the first name of the oldest overseas student
4. Write a script to count the number of students that are 30 years or older
5. Select the average age for students on a course. Show the course id and course name with this.
6. What is the total amount of fees being paid by all the students
7. What is the total amount of fees being paid by overseas students
8. Write a query that shows which course is the most profitable.
9. Modify the above query and use the ‘having’
10. What is the average amount of fees paid by people over the age of 20?
11. What fee is paid by the youngest student?

4.4 Private Study Exercises

You should spend approximately 7.5 hours on the Private Study for this topic. You should use this time to complete the exercises below as directed by your lecturer and to review the contents of this topic.

Exercise 1:

Read all the relevant material for this topic. This should including the slides and slide notes along with the sections from at least one relevant textbook.

You should prepare a PowerPoint presentation or discussion document that examines the following issues:

- The Purpose of SQL
- The difference between Data Definition Language (DDL) and Data Manipulation Language (DML)
- Parts of the Data Manipulation Language
- The Structure of a Select statement
- The function of the aggregate functions

Be prepared to present and discuss your ideas in the tutorial session.

Exercise 2:

You have been asked to develop a further set of possible transactions for the Marlowe Interior system. List and give details of possible transactions that would be likely for this system.

You should think about aspects of the business that will need to be catered for by the new system. Marlowe Interiors should be able to keep data on their customers and add new customers if they need to. Customers' details might also change. The main part of their business involves keeping track of jobs, who is working on them and what parts are used. With all parts of their business there might be new data such as new workers, parts of suppliers. Marlowe Interiors also needs to generate various paper-based reports such as job sheets and invoices.

Think about these aspects of the business and others you have identified. While looking at the ER diagram create a list of transactions that support these activities.

4.5 Tutorial Notes

The tutorials for this topic will last for 2 hours. You can expect to spend some of this time discussing your answers to the Private Study exercises with your lecturer and other students. Your lecturer will direct you on completing the tasks below.

Exercise 1:

In small groups, discuss your findings from Private Study Exercise 1 and develop a group presentation that gives an overview of the issues mentioned in the exercise.

Each group should present their ideas and the information should be discussed as a whole class group.

Exercise 2:

Discuss Private Study Exercise 2 as a whole class group.

Exercise 3:

a. Outline the ways in which one-to-one relationships can be represented in a database.
b. Explain what is meant by the term ‘recursive relationship.’
c. Why is it important to resolve many-to-many relationships?

Exercise 4:

Define a CRUD matrix for the new transactions that you have identified for Marlowe Interiors.

Define data types and domains for Marlowe Interiors system.

Topic 5: Data Retrieval 2

5.1 Learning Objectives

This topic provides an overview of data retrieval operates in more complex situations. On completion of the topic, you will be able to:

- Outline the concept of referential integrity and say why it is important in a relational database;
- Understand how to retrieve data from one or more tables using join;
- Understand how to retrieve data from one or more tables using sub-queries.

5.2 Timings

Lectures:	2 hours
Laboratory Sessions:	1 hour
Tutorials:	2 hours
Private Study:	7.5 hours

5.3 Laboratory Sessions

The laboratory time allocation for this topic is 1 hour. These exercises use the Students and Courses tables that were created in the laboratory session for Topic 1.

Exercise 1:

To investigate some of the joins we need to insert some new data.

1. Insert a new student into the students table.

ID	6
First_name	Arno
Last_name	Laski
Student_type	Overseas Student
Course_id	NULL
Age	25

Note that you will need to leave the course_id null.

2. Insert the following new course: course id = 4, course_name = 'Art'

Exercise 2:

Execute the following query:

```
select course_name, student_id, first_name
from students, courses
where students.course_id = courses.course_id;
```

Note that it does not bring back any of the new data that you have entered. This is because the basic join used here excludes any unmatched columns. To retrieve the unmatched columns then an outer join should be used.

1. Re-write the above query so that it shows even students who are not yet allocated to a course.

2. Write a query that retrieves the course_name and the students' first and last names and includes courses that do not yet have any students allocated to them.

3. To get both the courses without students and the students without courses requires the use of a full outer join. However this features is not always supported in implementations of SQL. One way of doing it is to use the UNION operator to join the two previously written queries. Write the query using the method available from your particular version of SQL.

4. Select all the students who are over the age of 20. Include those students who are not yet allocated a department

5. Now select all the students who are over 20 who are paying more than 6000 in fees. This will require the addition of a sub-query to the query in the previous question.

6. Select the course name and average age of students on that course. Include students who do not have a course.

5.4 Private Study Exercises

You should spend approximately 7.5 hours on the Private Study for this topic. You should use this time to complete the exercises below as directed by your lecturer and to review the contents of this topic.

Exercise 1:

Read all the relevant material for this topic. This should including the slides and slide notes along with the sections from at least one relevant textbook.

You should prepare a PowerPoint presentation or discussion document that examines the following topics with the use of examples:

- The necessity of joining tables.
- Simple joins
- Multi-table joins
- Types of Outer Joins

Be prepared to discuss this in the tutorial session.

Exercise 2:

In a fully developed database system, data is often accessed and manipulated through programs known as applications. These might be reports, screen based forms, web-pages or other types of tool.

Using the resources available to you, investigate the types of applications available for a database system. Prepare a discussion document about them.

Be prepared to discuss this in the tutorial session.

Exercise 3:

Using Web resources, investigate the typical application software available for one of the following DBMS products and create a document that details this:

- Oracle
- MySQL
- MS SQL-Server

The format of your document is up to you, it could be a text based document, mind-map or other illustration.

Be prepared to discuss this in the tutorial session.

5.5 Tutorial Notes

The tutorials for this topic will last for 2 hours. You can expect to spend some of this time discussing your answers to the Private Study exercises with your lecturer and other students. Your lecturer will direct you on completing the tasks below.

Exercise 1:

In small groups, discuss your findings from Private Study Exercise 1 and develop a group presentation that gives an overview of the topics mentioned in the exercise.

Each group should present their ideas and the information should be discussed as a whole class group.

Exercise 2:

In small groups, discuss your findings from Private Study Exercise 2 and develop a group presentation that gives an overview of the types of applications available for database system.

Each group should present their ideas and the information should be discussed as a whole class group.

Exercise 3:

In small groups, discuss your findings from Private Study Exercise 3 and develop a group presentation that gives an overview of the types of applications available for database system.

Each group should present their ideas and the information should be discussed as a whole class group.

Exercise 4:

a. Give a definition of referential integrity and say why it is important.
b. What role do foreign keys play in the construction of queries?
c. What is the difference between a sub-query and a join?
d. What is a correlated sub-query?

Exercise 5:

Define the joins and scripts for all transactions that have been defined for Marlowe Interiors.

Topic 6: Physical Design 1

6.1 Learning Objectives

This topic provides an overview of physical design. On completion of the topic, you will be able to:

- Understand the purpose of physical design;
- Map a logical database design to a physical database design;
- Design tables for the chosen database product.

6.2 Timings

Lectures:	2 hours
Laboratory Sessions:	1 hour
Tutorials:	2 hours
Private Study:	7.5 hours

6.3 Laboratory Sessions

The laboratory time allocation for this topic is 1 hour.

Exercise 1:

You should create the tables that you have developed as part of the Marlowe Interiors example you have been working on in the previous tutorial.

6.4 Private Study Exercises

You should spend approximately 7.5 hours on the Private Study for this topic. You should use this time to complete the exercises below as directed by your lecturer and to review the contents of this topic.

Exercise 1:

Read all the relevant material for this topic. This should including the slides and slide notes along with the sections from at least one relevant textbook.

You should prepare a PowerPoint presentation or discussion document that gives an overview of the following topics:

- The Purpose of Physical Design
- The Knowledge needed of chosen DBMS
- Steps 3 to 8 of design process (i.e. the steps of Physical Design).

Exercise 2:

For each of the 3 areas discussed above, devise 5 quiz questions (so you should have a total of 15 questions), along with the answers to these questions. These will be used in the tutorial to conduct a peer to peer quiz based on the topics studied.

Exercise 3:

Provide a documented mapping of the logical database design to the physical design for Marlowe interiors.

All tables and possible domains should now be defined:

a. Produce a table design document. This could be a modified ER document.
b. Produce a document of each table in the format shown in the slides.

6.5 Tutorial Notes

The tutorials for this topic will last for 2 hours. You can expect to spend some of this time discussing your answers to the Private Study exercises with your lecturer and other students. Your lecturer will direct you on completing the tasks below.

Exercise 1:

In small groups, discuss your findings from Private Study Exercise 1 and develop a group presentation that gives an overview of the topics mentioned in the exercise.

Each group should present their ideas and the information should be discussed as a whole class group.

Exercise 2

In small groups, using the questions that you wrote for Private Study Exercise 2, select the five best questions under each heading from your group. Each group should take it in turns to ask the rest of the class their questions.

Exercise 3:

In small groups, discuss your findings from Private Study Exercise 3.

Each group should present their ideas and the information should be discussed as a whole class group.

Exercise 4:

a. What is the difference between a data type and a domain?
b. Why is it important to have a knowledge of your chosen DBMS when performing physical design?
c. How can CASE tools aid the physical design process?
d. Why is it useful to be able to alter a table using SQL after it has been created?

Exercise 5:

Write the SQL create scripts for each of the tables in Marlowe Interiors.

Topic 7: Physical Design 2

7.1 Learning Objectives

This topic provides an overview of derived data. On completion of the topic, you will be able to:

- Understand the concept of derived data
- Design a representation of derived data
- Recognise the trade-offs between different ways of implementing derived data

7.2 Timings

Lectures:	2 hours
Laboratory Sessions:	1 hour
Tutorials:	2 hours
Private Study:	7.5 hours

7.3 Laboratory Sessions

The laboratory time allocation for this topic is 1 hour.

Exercise 1: Inserting data into the Marlowe Interiors database

Data should be inserted into the database for Marlowe Interiors. Referenced tables will need to be populated first. Students should insert a representative sample of data for example about 10 customers, 10 jobs etc.

Please note that multiple values can be inserted in the format:

```
insert into courses values
(5,'Biology'),
(6,'Psycology'),
(7,'Music')
```

Where data is inserted into a table where there is an auto increment or sequence on the primary key then the other rows should be specified and inserted and the primary key will be automatically generated.

```
insert into workers (worker_first_name, worker_last_name, worker_type_code)
values ('Fred','Smith','DEC')
```

Exercise 2: Practising comparison operators

SQL has a number of comparison operators that lets the user perform takes where a simple equality statement (“=”) will not do.

- LIKE – where the condition is that the value is similar to the specified condition
- IN – where a column value is equal to any of the values that are set out in a list supplied in the query
- BEWEEEN … AND – specifies a range of values
- IS NULL – specifies that a condition where the rows retrieve has a null value in that column

Run this example against the student table you created in the first workshop.

```
select * from students
where first_name like 'D%'
```

This example retrieves the student details where a first name begins with the letter ‘D’. Note the use of the wild card operator ‘%’: this is used to mean any value.

See what happens If you try to do the following using the equality operator:

```
select * from students
where first_name = 'D%'
```

Now use the tables from the first workshop to complete the following exercises.

Exercise 3:

1. Select all the students whose last name begins with ‘S’.
2. Select all the students whose last name begins with ‘S’ and whose first name begins with ‘H’
3. The format of the BETWEEN statement is of the form:

   Where <column_name> Between <first_value> And <second_value>

   Use this to select all the students who are between the ages of 25 and 30.
4. The format of the IS NULL statement is to simply write ‘is null’ in the where condition.

   Use this to select the first name and last name of all overseas students between 19 and 25 who do not have a course (i.e. their course_id will be null).
5. Use the IN operator to select course information from the courses table that you created in the lab session for Topic 1, but only for a set of specific courses.
6. Now, using this structure and joining with the student table, select the first and last names, and the course name, of all the students who are studying History or Geography.
7. Modify the last statement to select the first name and last name, and the course name for all students studying History and Geography whose first name starts with the letter ‘H’.

7.4 Private Study Exercises

You should spend approximately 7.5 hours on the Private Study for this topic. You should use this time to complete the exercises below as directed by your lecturer and to review the contents of this topic.

Exercise 1:

Read all the relevant material for this topic. This should including the slides and slide notes along with the sections from at least one relevant textbook.

You should prepare a PowerPoint presentation or discussion document that gives an overview of the following topics:

- What is derived data?
- The type of documents and applications which use derived data.
- Storing derived data as a database field or view vs. deriving at runtime.

Be prepared to discuss this in the tutorial session.

Exercise 2:

Collect as many documents as you can that contain fields that show examples of derived data. These could be receipts, order forms, application forms, invoices or customer order forms.

For a selection of at least 3 of these documents you should identify:

- Derived attributes
- The source of the derived attribute (i.e. in what other attributes)
- The method that has been used to derive any derived attribute. (i.e. what operations are performed to come up with the result).

7.5 Tutorial Notes

The tutorials for this topic will last for 2 hours. You can expect to spend some of this time discussing your answers to the Private Study exercises with your lecturer and other students. Your lecturer will direct you on completing the tasks below.

Exercise 1:

In small groups, discuss your findings from Private Study Exercise 1 and develop a group presentation that gives an overview of the topics mentioned in the exercise.

Each group should present their ideas and the information should be discussed as a whole class group.

Exercise 2:

In small groups, discuss the documents that you found when you completed Private Study Exercise 2.

This should then be discussed as a whole-class group.

Exercise 3:

a. Give an explanation of derived data and explain the different ways it can be represented in a database.
b. What are trade-offs for the different ways of representing derived data?
c. What types of aggregate functions can be used in deriving data?
d. Explain how non-numeric derived data could be used in enforcing rule in a database.

Exercise 4:

Design and specify the requirements for at least two derived attributes in the Marlowe Interiors system.

Topic 8: Physical Design 3

8.1 Learning Objectives

This topic provides an overview of database constraints. On completion of the topic, you will be able to:

- Define different types of constraints;
- Understand how to design and implement constraints on their chosen DBMS.

8.2 Timings

Lectures:	2 hours
Laboratory Sessions:	1 hour
Tutorials:	2 hours
Private Study:	7.5 hours

8.3 Laboratory Sessions

The laboratory time allocation for this topic is 1 hour.

Exercise 1:

1. Setting a date automatically and working out the length of a job. A date can be generated automatically. This is an example in from MySQL:

```
insert into jobs (start_date, job_type_code)
values (current_timestamp,'KIT')
```

Now set the end date to sometime in the future:

```
update jobs
set end_date = ('2012-7-04')
where job_id = 1
```

Using the datediff function in MySQL (or similar functions in other vendors versions of SQL) it is possible to derive the length of a job in a select statement:

```
select job_id, datediff (end_date, start_date) job_length
from jobs
where job_id = 1;
```

2. How much a job costs. The building blocks. Insert some values into the worker_jobs table.

```
Insert into worker_jobs values (1,1,'2010-01-01','2011-01-01')
```

The next query joins the worker_jobs table with the workers table and then the worker_type table.

It works out how many days have been spent on the job by doing the date difference function. It then multiplies this by the hourly rate and then by the number of hours spent per day on a job (in this case 12). It makes many assumptions (such as weekend working!) but it is the beginning of the building blocks for deriving data for producing a full invoice for a job, for example.

```
select worker_jobs.worker_id, workers.worker_type_code, worker_types.hourly_rate,
worker_jobs.start_date, worker_jobs.end_date, datediff(worker_jobs.end_date,
worker_jobs.start_date) days_on_job, datediff(worker_jobs.end_date,
worker_jobs.start_date)

* 12 * hourly_rate labour_cost
from worker_jobs, worker_types, workers
where worker_jobs.worker_id = workers.worker_id
and workers.worker_type_code = worker_types.worker_type_code;
```

Try to get this query working using the data and structures in your own version of the Marlowe Interiors database.

3. Attempt to write a similar query for working out the cost of parts used on a job.

8.4 Private Study Exercises

You should spend approximately 7.5 hours on the Private Study for this topic. You should use this time to complete the exercises below as directed by your lecturer and to review the contents of this topic.

Exercise 1:

Read all the relevant material for this topic. This should including the slides and slide notes along with the sections from at least one relevant textbook.

Your lecturer should have divided the class into small groups. In your groups you should prepare quiz questions based on the topics listed below. You should also provide a set of answers. The questions and answers should be supplied to your lecturer prior to the tutorial session.

Try to aim for at least two questions per bullet point as outlined below:

a. Rules

- Entity Integrity
- Referential Integrity
- Propagation constraints
- Domain constraints
- Table constraints

b. Types of propagation constraint:

- No action
- Cascade
- Set Default
- Set Null

c. Ways of enforcing constraints.

- As rules on tables
- Using the alter table
- Check constraints
- Domains

8.5 Tutorial Notes

The tutorials for this topic will last for 2 hours. You can expect to spend some of this time discussing your answers to the Private Study exercises with your lecturer and other students. Your lecturer will then direct you on completing the tasks below.

Exercise 1:

Your lecturer will run a class quiz based on the questions produced as part of Private Study Exercise 1.

Exercise 2:

a. What are the different ways of enforcing a domain constraint in a database management system?
b. What are the ways a business rule can be enforced in a database management system?
c. What are database triggers?
d. What is a sequence in a database system and what is it used for?

Exercise 3:

Define all the constraints for the Marlowe Interiors system.

Exercise 4:

Outline which propagation rules would be appropriate for the Marlowe Interiors database.

Topic 9: Physical Design 4

9.1 Learning Objectives

This topic provides an overview of transactions and related issues. On completion of the topic, you will be able to:

- Define transaction use
- Understand the concept of de-normalisation
- Understand the use of indexes
- Estimate the size of a database

9.2 Timings

Lectures:	2 hours
Laboratory Sessions:	1 hour
Tutorials:	2 hours
Private Study:	7.5 hours

9.3 Laboratory Sessions

The laboratory time allocation for this topic is 1 hour.

Exercise 1:

Any foreign key constraints that were not enforced earlier should now be implemented.

Note that in the past there have been problems with creating check constraints in MySQL where they have been created but then ignored.

Implementing check constraints to enforce business rules.

1. An example of a business rule might be that if a quantity of parts is inserted into the database then it should be greater than 0.

```
alter table job_parts
add check (quantity > 0)
```

2. More complex constraint to enforce a domain:

```
alter table customers
add constraint check (address_line2 in ('London','Birmingham','Manchester'))
```

3. More complex constraints can be added by using triggers. These are specific to individual vendors.

You should add any constraints that you have defined for the Marlowe Interiors database.

Exercise 2: Queries on the Marlowe Interiors database

1. Create a query that shows all the jobs of a particular type
2. Create a query that shows customers who have more than one job
3. Create a query that shows which jobs a particular part is used on
4. Create a query that shows all the parts supplied by a particular supplier
5. Show all the jobs in a particular region

9.4 Private Study Exercises

You should spend approximately 7.5 hours on the Private Study for this topic. You should use this time to complete the exercises below as directed by your lecturer and to review the contents of this topic.

Exercise 1:

Read all the relevant material for this topic. This should include the slides and slide notes along with the sections from at least one relevant textbook.

You should prepare a PowerPoint presentation or discussion document that gives an overview of what a transaction is and what the basic operational units of a transaction are.

Be prepared to discuss this in the tutorial session.

Exercise 2:

Use a copy of the table diagram and, with valid assumptions, estimate the size of the Marlowe Interiors database.

Be prepared to discuss this in the tutorial session.

9.5 Tutorial Notes

The tutorials for this topic will last for 2 hours. You can expect to spend some of this time discussing your answers to the Private Study exercises with your lecturer and other students. Your lecturer will direct you on completing the tasks below.

Exercise 1:

In small groups, discuss your findings from Private Study Exercise 1 and develop a group presentation that gives an overview of the topics mentioned in the exercise.

Each group should present their ideas and the information should be discussed as a whole class group.

Exercise 2:

In small groups, discuss your estimates of database sizing. Check the relationships to ensure consistency between entities. For example where the average ratio is used this should act as a multiplier of rows between relevant entities as illustrated in the slide show example.

Exercise 3:

a. What is meant by the term ‘performance’ with regard to database systems?
b. What is an index and what types of index are available?
c. What are the advantages of using indexes?
d. What are the potential problems with using indexes?

Exercise 4:

Your tutor will supply you with some blank transaction analysis forms either on paper or electronic format. One copy is provided below

All transactions previously identified as part of the Marlowe Interiors system should be fully documented now using transaction analysis forms for each.

Transaction Analysis Form **Date:** **Transaction Reference and Name:** **Transaction volume** **Average:** **Peak:**
<Insert query to be used here>
Access **Entity** **Type of Access** **Average Number** **Peak Number**

Topic 10: Distributed Databases

10.1 Learning Objectives

This topic provides an overview of distributed databases. On completion of the topic, you will be able to:

- Recognise the need for distributed data
- Define the main features of a distributed database
- Define the different types of distributed databases

10.2 Timings

Lectures:	2 hours
Laboratory Sessions:	1 hour
Tutorials:	2 hours
Private Study:	7.5 hours

10.3 Laboratory Sessions

The laboratory time allocation for this topic is 1 hour.

Exercise 1:

1. The syntax for creating indexes:

```
create index cust_id_index on customers(customer_id)
```

This creates an index called cust_id on the customers_id column on the customers table.

To create a unique index on 2 columns:

```
Create unique index cust_name_idx ON
customers(customer_first_name,customer_last_name);
```

In this case it might not be a good idea to do this because it is a very real possibility that Marlowe Interiors could have customers with the same name.

So drop the index:

```
drop index cust_name_idx on customers
and recreate as a non-unique index
create index cust_name_idx ON
customers(customer_first_name,customer_last_name);
```

2. Now create appropriate indexes for Marlowe Interiors.

Views are created by nesting an SQL statement inside the create view statement as follows:

```
create view london_customers as
(select * from customers
where address_line2 = 'London');
```

This creates a view called ‘london_customers’ which will contain all customers who have London in address_line2.

Views could implement some of the derived data from earlier.

3. Create a view based on the query you developed earlier which showed the total cost of parts for each job.

4. Create a view based on the query used to calculate labour.

10.4 Private Study Exercises

You should spend approximately 7.5 hours on the Private Study for this topic. You should use this time to complete the exercises below as directed by your lecturer and to review the contents of this topic.

Exercise 1:

Read all the relevant material for this topic. This should including the slides and slide notes along with the sections from at least one relevant textbook.

You should prepare a PowerPoint presentation or discussion document that gives a **detailed** review of the following topics:

- Definition of distributed databases
- Why distributed databases are needed
- Fragmentation
- Replication
- The Foundation Rule for Distributed Databases

Be prepared to discuss this in the tutorial session.

Exercise 2:

One part of expanding the use of databases in an organisation concerns the concept of scalability. This concerns the ability of a database to operate in the same way at different sizes. Size here refers to the number of rows in a database and the number of users.

Before any move to distributing data then questions of scalability should be dealt with.

Write an email to the head of Marlowe Interiors discussing the following issues of database scalability:

- The components of a large database
- Increasing database storage capacity
- Using memory to make the database faster
- Scaling up to more powerful CPU and servers
- Evolution of the commodity server
- Scaling out across multiple database servers
- High-speed connections between database components
- Database transaction processing
- Optimizing database design and configuration

You may find the following resources useful:

- This taxonomy is adapted from “Database Scalability” by Base One International Corporation (Available Online from www.boic.com/scalability.htm)

- Oracle: Karam, S. (undated) *Introduction to Oracle Scalability Issues.* [Available Online] http://www.dba-oracle.com/t_scalability_features.htm

- SQL Server: Kerwin, D. (undated) *Achieving massive scalability with SQL Server.* [Available Online] http://www.sql-server-performance.com/2003/massive-scalability/

10.5 Tutorial Notes

The tutorials for this topic will last for 2 hours. You can expect to spend some of this time discussing your answers to the Private Study exercises with your lecturer and other students. Your lecturer will direct you on completing the tasks below.

Exercise 1:

In small groups, discuss your findings from Private Study Exercise 1 and develop a group presentation that gives an overview of the topics mentioned in the exercise.

Each group should present their ideas and the information should be discussed as a whole class group.

Exercise 2:

a. What are vertical fragmentation and horizontal fragmentation? Give an example of each.
b. Outline the advantages of distributed databases for an organisation.
c. What are the potential disadvantages of distributed databases?
d. Outline the main features of both homogeneous and heterogeneous distributed databases.
e. What is a federated database?
f. What features should a distributed DBMS have?
g. How does designing a distributed database system affect the physical design process?

Exercise 3:

Marlowe Interiors want to open 2 new branches, one in Birmingham and one in Manchester. Each branch will have its own set of clients and its own employees. It will, however, share use of the same information about suppliers.

With the aid of a diagram outline how a distributed database could be implemented for Marlowe Interiors.

Topic 11: Data Warehouses

11.1 Learning Objectives

This topic provides an overview of data warehouses. On completion of the topic, you will be able to:

- Understand the potential need for a data warehouse;
- Differentiate between online transaction processing systems and data-warehouse system;
- Identify the main components of a data warehouse.

11.2 Timings

Lectures:	2 hours
Laboratory Sessions:	1 hour
Tutorials:	2 hours
Private Study:	7.5 hours

11.3 Laboratory Sessions

The laboratory time allocation for this topic is 1 hour.

Exercise 1: Queries and transactions for Marlowe Interiors

Use some of the views created earlier to complete the following.

1. Using the view you created in the previous exercise create a query that retrieves the total costs of parts, the total cost of labour and the total cost of the entire job.

2. Modify the query to include those jobs where the costs have not yet been calculated. You will need to use an inner join.

3. Now build on what you have learnt to create a full bill for a job. This should include the customer's name, the parts and labour used and the final totals.

11.4 Private Study Exercises

You should spend approximately 7.5 hours on the Private Study for this topic. You should use this time to complete the exercises below as directed by your lecturer and to review the contents of this topic.

Exercise 1:

Read all the relevant material for this topic. This should including the slides and slide notes along with the sections from at least one relevant textbook.

You should prepare a **detailed** review of the following topics:

- Why organisations might need a data warehouse
- The difference between OLTP and OLAP
- Why data designed for OLTP systems might not be suitable for OLAP systems
- How bringing data together can aid an organisation in its decision making

Be prepared to discuss this in the tutorial session.

Exercise 2: Management Support Systems

You have been asked to prepare an email or memo detailing the different sorts of management information systems and approaches to data analysis that will be able to utilise a proposed data warehouse.

Give an overview of:

- Decision Support Systems
- Management Information Systems
- Data Mining

Be prepared to discuss this in the tutorial session.

11.5 Tutorial Notes

The tutorials for this topic will last for 2 hours. You can expect to spend some of this time discussing your answers to the Private Study exercises with your lecturer and other students. Your lecturer will direct you on completing the tasks below.

Exercise 1:

In small groups, discuss your findings from Private Study Exercise 1 and develop a group presentation that gives an overview of the topics mentioned in the exercise.

Each group should present their ideas and the information should be discussed as a whole class group.

Exercise 2:

In small groups, discuss your findings from Private Study Exercise 2 and develop a group presentation that gives an overview of the topics mentioned in the exercise.

Each group should present their ideas and the information should be discussed as a whole class group.

Exercise 3:

a. Describe the four key features of a data warehouse: subject orientation; integration; time-variance; and non-volatility.
b. Within the data warehouse functional model what is the acquisition process and what activities take place within it?
c. Why is the definition of metadata important when building a data-warehouse?
d. Outline the steps needed when building a data warehouse.
e. Define the three main schemas used for data warehouses.
f. What are the four key features of OLAP?
g. Give an example of multi-dimensional data.

Exercise 4:

Marlowe Interiors has expanded rapidly and taken over several other building and interior design organisations. Each of these organisations has its own database built using a different vendor to that of Marlowe Interiors.

What are the challenges that would face the builder of a data warehouse for Marlowe Interiors? What benefits would Marlowe Interiors gain from building a data warehouse?

Topic 12: Module Overview

12.1 Learning Objectives

This topic provides an overview of the module as a whole. On completion of the topic, you will be able to:

- Summarise the key module topics
- Give an outline of the knowledge needed about each topic for assessment purposes

12.2 Timings

Lectures:	2 hours
Laboratory Sessions:	1 hour
Tutorials:	2 hours
Private Study:	7.5 hours

12.3 Laboratory Sessions

The laboratory time allocation for this topic is 1 hour. The following exercises introduce some further structures in SQL. Use the Students and Courses tables created in Topic 1.

Exercise 1: SQL Logical Operators

There are 3 logical operators in SQL.

- AND this joins two or more conditions together and specifies that both must be true. You will have been using this to compose joins throughout the exercises.
- OR this joins two or more conditions and specifies that at least one of them must be true
- NOT this specifies that the condition must not be true

Example

To get the students whose surname is surname is 'Smith' or 'Singh'

```
select first_name, last_name
from students
where last_name = 'Smith' or last_name = 'Singh'
```

Exercise 2:

1. Select students whose age is 19 , 20 or 21.

2. Select the first name, last name and course name for students whose age is 21, 20 or 19 and whose course is History or Geography.

3. The NOT operator negates a condition. It can be put in front of a condition so that the opposite is true.

4. Modify the query from 2. using the NOT operator to select the first name, last name and course name for students whose age is 21, 20 or 19 and whose course is NOT History or Geography.

Exercise 3: UNION, UNION ALL, INTERSECT

1. Create a table to store the details of students who are studying by distance learning.

2. Insert the following values into the table:

Student ID	First Name	Last Name	Course ID	Age
8	Gary	Smith	3	33
9	Charlie	Brown	3	21
10	Ray	Patel	2	21
11	Amelia	Drongo	2	21
3	Dawn	Cove	1	21

The UNION, UNION ALL and INTERSECT operators are set operations. They only work where a query specifies the same types of rows from the tables being used. For example to select the student id, first name and last name from the students on both the normal Students table, and those on the Distance_students table then use the following:

```
select student_id, first_name, last_name
from students
UNION
select student_id, first_name, last_name
from distance_students
```

Note that the student 'Dawn Cove' appears only once. That is the function of the UNION operator.

3. Now run the same query as above but instead of UNION use the UNION ALL operator.

4. Select all the student from Students and from Distance_Students who are under 30 years of age.

   The INTERSECT operator selects just those rows that appear in both tables.

   The LIMIT operator get extreme (highest or lowest) values.

   For example, to get the two oldest students from the student table:

```
select first_name, age
from students
order by age desc
limit 2;
```

5. Run this query.

6. How would you get the youngest 3 students? Run the necessary query.

12.4 Private Study Exercises

You should spend approximately 7.5 hours on the Private Study for this topic. You should use this time to complete the exercises below as directed by your lecturer and to review the contents of this topic.

Exercise 1:

These questions review important material covered in the module. Provide written answers for each question and be prepared to discuss them in the tutorial session:

1. What is meta-data and why is it important in a database system?
2. What is 'requirements gathering' and what role does it play in database development?
3. What are the key properties of a relation
4. Give a definition and example of the concept of functional dependence.
5. Give a definition of the concept of a domain and explain how it is related to the concept of a data-type.
6. What is the difference between a super key and a candidate key?
7. Explain the concept of a recursive relationship.
8. What is an aggregate function in SQL and why must it be used with the Group By clause?
9. What is the purpose of an Outer Join in SQL?
10. What is the Cartesian product of two tables in SQL?
11. What is the difference in purpose between logical and physical design?
12. What is a CASE tool and how might it be used in database development?
13. Define the concept of derived data.
14. How are SQL aggregate functions used in generating derived data?
15. What is the entity integrity rule and why is there a danger of it being compromised in some vendor's implementations of the relational model?
16. What is a sequence or auto increment used for?
17. What does the term 'performance' refer to with regard to database transactions?
18. What is an index on a database and what is its purpose?
19. What are the advantages of distributed databases?
20. What does the concept of 'scalability' refer to?

21. What is meant by the term ‘non-volatility’ with regard to data-warehouse systems?

22. Why is meta-data an important concept in the development of data-warehouses?

Exercise 2:

Look over the module materials, noting any areas where you require further explanation. Prepare a detailed set of notes on these areas and take them to the tutorial session.

12.5 Tutorial Notes

The tutorials for this topic will last for 2 hours. You can expect to spend some of this time discussing your answers to the Private Study exercises with your lecturer and other students. Your lecturer will then direct you on completing the tasks below.

Exercise 1:

In small groups discuss the areas of the module content that you do not fully understand. Attempt to answer each other's questions. If you are not able to answer these questions between you then make a note of the questions and raise these with your tutor.