B5.

a) Consider the concept of a distributed database and answer the questions below:

    i) Define the concept of 'distributed database'.

          (4 marks)

    ii) C.J. Date's criteria for distributed databases includes 'local site independence' and 'fragmentation independence'. Describe what these are and what they mean for DBAs and users respectively.

          (4 marks)

b) The Two Phase Commit protocol (2PC) enables transaction management for distributed relational database systems.

Explain how 2PC works in the context of a database spread over three RDBMSs (Relational Database Management System). Ensure that your answer covers the scenario where all RDBMSs are able to complete a transaction and the situation where at least one RDBMS is not able to commit.

You might wish to use a diagram as part of your answer.

          (7 marks)

c) Data replication refers to creating copies of data on additional database servers.

Explain the advantages and disadvantages of data replication listed below considering each in the context of a distributed multiplayer online game and a banking application.

    i) The advantages due to load reduction and high availability.

          (5 marks)

    ii) The disadvantages due to data loss and data inconsistency.

          (5 marks)

End of Examination

Section A
Answer Section A questions in Answer Book A

A1.

a) Optimistic and Pessimistic Concurrency Control use different strategies and techniques in order to achieve database integrity.

Compare the following concurrency control approaches and explain why each particular technique is used for that particular concurrency control approach. Give examples of any disadvantages in the use of each approach and technique.

Concurrency Control Approach	**Associated Technique**
Pessimistic.	Locking.
Optimistic.	Multiversion Concurrency Control.

(10 marks)

b) Briefly describe how databases can recover from the following types of failure:

   i) System failure (loss of entire machine/server - often due to power outage).

   ii) Catastrophic failure such as loss of entire installation/building.

(8 marks)

c) Using an example, explain what is meant by cascading rollback. Explain why dirty reads can result in a cascading rollback.

(**7 marks)**

d) Relational databases organise data in tables queried by SQL while RESTful WebServices typically exchange data via Json formats. Use code or pseudocode to describe the relevant part for a service that executes a query and converts the result to Json for the following table and Json output.

(7 marks)

Friends table:

name	age	gender
Steve	29	male
Sophie	27	female

JSON:
```
var friends = [
    {"name" : "Steve", "age" : "29", "gender" : "male"},
    {"name" : "Sophie", "age" : "27", "gender" : "female"},
];
```

Section B
Answer Section B questions in Answer Book B

B4.

a) Briefly describe the concept of 'reactive tuning' (also known as 'bottom-up tuning') in the context of database performance.

(2 marks)

b) Your DBA has identified that your database is not performing at its optimum and has concluded that it is likely that hardware limitations are impeding performance.

Discuss how CPU, RAM and Disks can create bottlenecks, how these manifest themselves and how you would resolve them.

(8 marks)

c) Role based access control in databases is one mechanism available to DBAs to ensure data confidentiality. Consider a payroll database with `employee` and `salary` tables, and user roles called `payroll_manager` and `senior_HR_advisor`. Answer the following:

i) Define data confidentiality.

(2 marks)

ii) Explain how the SQL `GRANT` statement supports a role based approach to security control.

(2 marks)

iii) Provide the SQL statements to:

1. Allow only `payroll_managers` to query the `salary` table.

2. Explicitly prevent `senior_HR_advisors` from updating the `birthdate` column in the `employee` table.

(4 marks)

A2.

Consider Fig A2 below, which represents an Entity Relationship (ER) data model used in an order processing application. Assume that all orders contain at least one product and that there may be products that have never been ordered.

a) Using a standard Object Oriented modelling notation, such as a UML Class model, transform the ER model (Fig A2) to an Object Oriented model that accommodates a representation of the following modelling constructs:

- Classes.
- The association (or relationship) between classes.
- The cardinality of each association (either 0..1, 1:1 0..N, 1,,N).
- One example of generalisation/specialisation (inheritance) between classes.

(10 marks)

Fig A2 Entity Relationship model
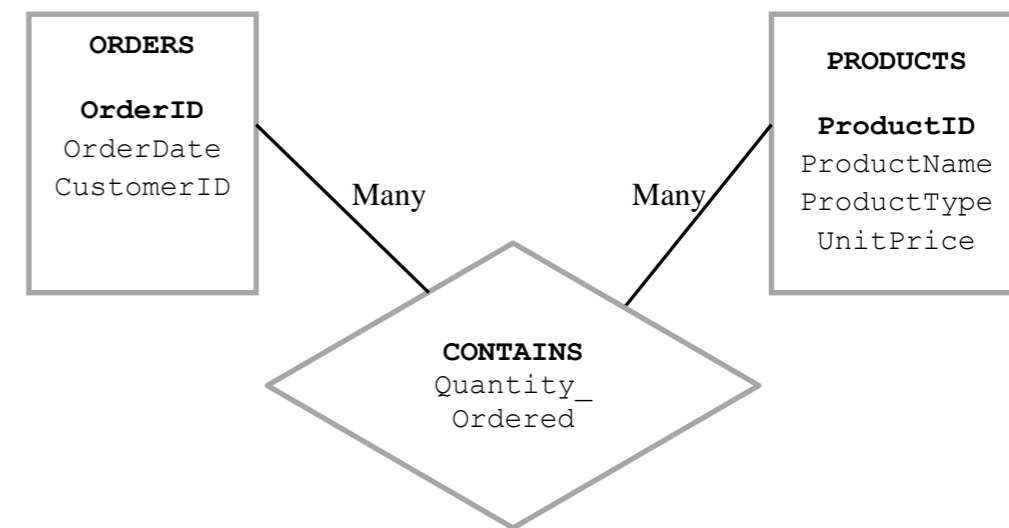
b) The Object-Oriented Programming (OOP) paradigm is a method of programming based on a hierarchy of classes, and well-defined and cooperating objects. In OOP, the software objects are used to model real-life ones, such as cars or employees.

i) Briefly describe the so called 'impedance mismatch' that occurs when interfacing relational databases with the OOP paradigm.

(5 marks)

ii) Describe methods and techniques that are used to integrate OOP with a relational database system.

(10 marks)

[Turn Over]

A3.

The UK population census is a vital resource of statistical information used in social scientific research and policy development, providing a snapshot of demographic and social life in the UK that helps inform government policy at local and central level.

A database stores census data collected from around 40 million households every 10 years.

A data warehouse is required that will store census data that has been transformed and aggregated into statistical data, for example, area counts. The data is usually of individuals or households with particular characteristics.

a)   Explain how a data warehouse differs in content, structure and function from an Online Transaction Processing (OLTP) database.

Use the following points to structure your answer:

- The type of data and characteristics of data in a data warehouse.
- The organisation/design of the data.
- The types of transactions and queries operating on the data.

(10 marks)

b)   Briefly describe the CUBE and ROLLUP operators that have been included in the SQL-99 standard. Support your answer with examples relating to the census data scenario above.

(6 marks)

c)   Describe the job roles and principal duties of the following personnel needed to support the development of a data warehouse solution for handling census data.

Include in your answer a key technique that supports **each** role.

- Business Intelligence (BI) Data Analyst.
- Data Architect.
- Extract Transform Load (ETL) lead developer.

(9 marks)

[Turn Over]