B5.

Keebo is a large company that sells a range of products and services. It has sites in many different countries where most of its products for sale are sourced or manufactured.

Keebo wants to build a large central repository to gather information about its business activities and processes from all of its sites. Orders and sales transactions made in the past are recorded in a range of databases and other data stores located at most of its sites.

Use the above scenario as a source of examples when answering the following question parts. State any assumptions you need to make.

a) Explain the benefits that Keebo would gain, and the challenges they may encounter, from using a Data Warehouse.

(8 marks)

b) Explain briefly what is meant by ETL and explain why ETL is an essential process in the development of a Data Warehouse for Keebo.

(5 marks)

c) With the aid of a suitably annotated diagram, explain how the following data modelling concepts are applied to a Snowflake Schema.

- Facts
- Dimensions and Attributes
- Multi-Dimensional modelling.

(12 marks)

End of Examination

This page is left intentionally blank.

B4.

Consider the following diagram (fig A1) that depicts a standard n-tier web database software framework.

Fig A1. n-tier web database software architecture.
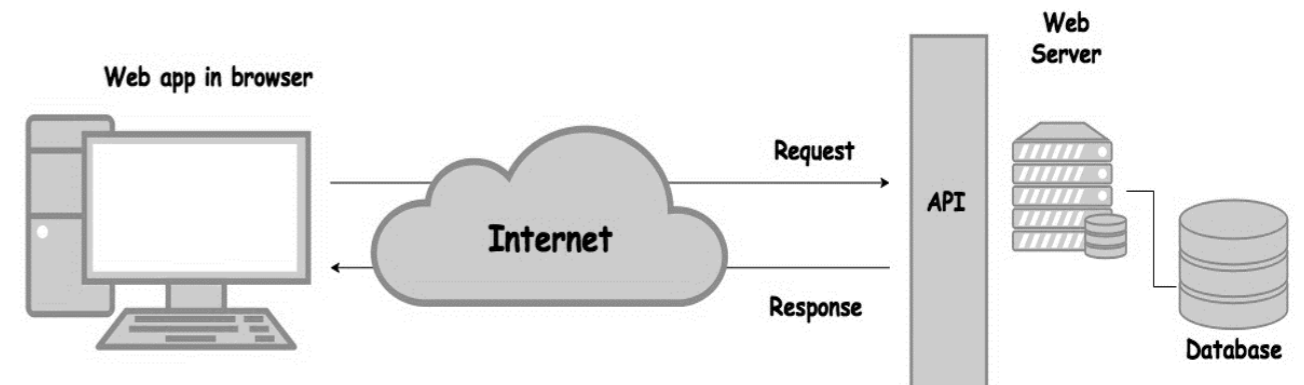
a) Explain how a request for data is handled in the software framework given in Fig A1 above. Include in your answer an explanation of the function of a Web App; API (Application Programming Interface) and the Web Server.

(9 marks)

b) Web services can be built on top of the software framework depicted in Fig A1.

Describe the main and distinctive characteristics of a Web Services approach based on XML/SOAP compared with the approach depicted in Fig A1.

Your answer should include reference to various components such as SOAP, XML, WSDL and UDDI that facilitate seamless interchange of data.

(8 marks)

c) JSON (JavaScript Object Notation) is an alternative data format to XML for transporting data over the internet.

Compare and contrast JSON with XML in terms of how they model and represent data and discuss the relative merits of each of them.

(8 marks)

[Turn Over]

A3.

a) Database systems need to operate reliably both in the context of failure and concurrent access by various users. In this context consider the following questions.

    i) Using your own example and simple diagrams, explain the concept of eventual data consistency via BASE (Basic Availability, Soft-state, Eventual consistency) principles**.**

            (5 marks)

    ii) Using your own example, explain the concept of serialisability in the context of transactions.

            (5 marks)

    iii) Using your own examples, explain the concepts of row level locking and table level locking.

            (5 marks)

b) Your database uses Deferred Update when writing data to disk.

    i) Briefly describe what this means for your database in operation.

            (4 marks)

    ii) Consider the following log and describe what needs to be done for each of the transactions `T1`, `T2` and `T3` in the case of a server crash just after `LogID = 10` is created.

            (4 marks)

    iii) If Log entry `LogID = 8` would show an **ABORT** operation rather than **COMMIT** for `T1`, what would be the correct actions at the time of the crash after the abort for `T1`?

            (2 marks)

LogID	TxID	Time	Operation	Object
1	T1	8:15	START	
2	T1	8:16	UPDATE	Account 122
3	T2	8:16	START	
4	T2	8:19	INSERT	Account 148
5	T2	8:20	DELETE	Account 153
6	T2	8:21	UPDATE	Customer 211
7	T3	8:21	START	
8	T1	8:21	COMMIT	
9	T2	8:24	COMMIT	
10	T3	8:25	INSERT	Customer 231

A1.

Forbes and other press reported "Capital One Says Hacker Breached Accounts Of 100 Million People; Ex-Amazon Employee Arrested". This is one of many examples of database security having failed.

a) How would you embed database security at the following development stages in order to reduce the risk of security breaches?

    i) Database requirements

            (3 marks)

    ii) Database design.

            (3 marks)

b) Explain the following advanced Database Management System features and describe how they can be used to increase database security. Provide a simple code example for **each** of the two features showing its basic structure.

    i) Database Triggers

            (7 marks)

    ii) Stored Procedures.

            (6 marks)

c) Briefly explain the threats to online databases and detail how you would mitigate against the identified risks**.**

            (6 marks)

A2.

a) Consider the following tables from an order management system and the two SQL queries. Answer the question below.

```
Product(ProductID, Name, Description, Price)
OrderItem(Id, OrderID, ProductID, Quantity)
```

Note that `OrderItem.ProductID` is a foreign key to the `ProductID` column in the `Product` table.

Q1:
```
SELECT Name
FROM Product
    JOIN OrderItem
    ON Product.ProductID = OrderItem.ProductID
WHERE Quantity > 100
```

Q2:
```
SELECT Name
FROM Product
    JOIN
        (SELECT ProductID
         FROM OrderItem
         WHERE Quantity > 100) v1
    ON Product.ProductID = v1.ProductID
```

You have 100 rows in your Product table and 10000 rows in your OrderItem table and that 10% of the sales are for more than 100 units of a product. Your DBMS does not apply any query optimisation.

i) Which query is more efficient?

(1 marks)

ii) Justify your answer by explaining the execution of the query paying attention to the number of results returned in the various steps.

(4 marks)

b) Consider the following tables:

```
Customers (CID, name, address)
Wines (WID, name, producer, origin, grape, year)
Orders (OID, CID, WID, date)
```

And the following query:

```
SELECT Customers.address
FROM Customer
    JOIN Orders
        ON Customers.CID = Orders.CID
    JOIN Wines
        ON Orders.WID = Wines.WID
WHERE Wines.origin = 'Chile'
AND Wines.grape = 'Pinot Noir';
```

i) Draw a query tree that corresponds to the most efficient way of processing this query.

(12 marks)

ii) Assume there is a B-Tree index on the column "grape" of the table "Wines".

Also, assume that "grape" is declared as NOT NULL. For each of the following queries, explain how this index could be used when executing each query.

QueryX   `SELECT * FROM Wines WHERE grape = 'Pinot Noir'`

(2 marks)

QueryY   `SELECT COUNT(grape) FROM Wines`

(3 marks)

QueryZ   `SELECT * FROM Wines ORDER BY grape`

(3 marks)

[Turn Over]