

IMIS HIGHER DIPLOMA QUALIFICATIONS

**Database Development
(H4)**

WITH ANSWERS

DURATION: 3 HOURS

Candidates should answer ALL the questions in Part A and THREE of the five questions in Part B. Part A carries 40% of the marks available and Part B carries 60%. Candidates should allocate their time accordingly.

No reference material of any kind may be taken into the examination.

[Turn over]

PART A. Answer ALL questions in this section**Question A1.**

Explain the following advantages and disadvantages as they affect an organisation considering implementing a database Management System (DBMS).

Advantages

- More information for the same amount of data – *sharing of data across functional areas*
- Improved data integrity – *DBA can define integrity constraints in the schema and DBMS will enforce them*
- Increased concurrency – *DBMS will handle multiple users access the same data at “the same time”*

Disadvantages

- Cost of conversion – *there may be a need for extra hardware, staff training file conversion etc.*
- Greater impact of failure – *all users and applications rely on the centralised DB so unavailability potentially impact all of them*

1 mark for each. Explanation must include a suitable example

5 marks

Question A2.

Explain the meaning of data independence, particularly logical data independence and physical data independence.

1 mark for explaining the concept of schemas or views

2 marks for logical independence – changes in the conceptual schema (logical view) do not affect the external schema (the user views)

2 marks for physical independence – changes in the internal schema (the physical definition) do not affect the conceptual schema.

5 marks

Question A3.

Before embarking on designing a database solution to a problem it is advisable for a designer to define the scope and boundaries. Explain what is meant by this giving suitable examples.

Scope – extent of the design according to operational requirements e.g. whole organisation or just some departments – 2 marks

Boundary – external limits e.g. hardware/software limitations, time/cost limitations – 3 marks

5 marks

Question A4.

Explain the difference between the two SQL statements, showing how the insertion of a record might be affected

```
CREATE INDEX INDEX1 ON PRODUCT(PRODUCTNAME)
```

```
CREATE UNIQUE INDEX INDEX2 ON PRODUCT(PRODUCTNAME)
```

Create Index – a search on product name will be quicker because the index can be searched and used to point to the appropriate record(s) – there may be more than one product with the same name – 2 marks

Create Unique Index – as well as speeding up searching as before it also ensure that no records can be included that duplicate the product name – 3 marks

5 marks

Question A5.

Explain the function and purpose of a subquery in SQL, giving a suitable example to help your explanation.

A subquery is a query that is embedded (nested) inside another query – 2 marks

Suitable example – 3 marks

5 marks

Question A6.

Explain the difference between distributed databases and distributed processing.

Distributed Database is a logical (1) collection of data physically (1) distributed over a computer network (1) whereas Distributed Processing is a centralised (1) database that can be accessed (1) over the network

5 marks

Question A7.

Consider the following SQL table definitions for customers and their accounts at a bank.

CREATE TABLE CUSTOMER (

CUSTNUM CHAR (5) NOT NULL UNIQUE,

CUSTFNAME VARCHAR (20) NOT NULL,

CUSTLNAME VARCHAR (20) NOT NULL,

CUSTPHONE CHAR (12))

PRIMARY KEY CUSTNUM;

CREATE TABLE ACCOUNT (

CUSTNUM CHAR (5) NOT NULL,

ACCNUM CHAR (5) NOT NULL,

TYPE CHAR (1) DEFAULT "C",

BALANCE NUMBER (5,2))

PRIMARY KEY (ACCNUM)

FOREIGN KEY (CUSTNUM) REFERENCES CUSTOMER;

The CUSTOMER table is populated with the following data

CUSTOMER			
CUSTNUM	CUSTFNAME	CUSTLNAME	CUSTPHONE
1	Arnold	Nueba	044571234567
2	Charles	Mwamba	066778882341
3	Blessing	Nwere	998776512345

© Institute for the Management of Information Systems –

For each of the following INSERT statements for the ACCOUNT table, state whether the insert would be successful. If it would be successful show what the inserted record would look like. If it would not be successful explain why not.

a) INSERT INTO ACCOUNT VALUES (1, 20, ,55.05);

The INSERT would work 1 mark

ACCOUNT			
CUSTNUM	ACCNUM	TYPE	BALANCE
1	20	C	55.05

1 mark

b) INSERT INTO ACCOUNT VALUES (4, 21, "I", 45.60);

The INSERT would not work because the foreign key (CUSTNUM) does not exist in the CUSTOMER table

c) INSERT INTO ACCOUNT VALUES (3, 22, R, 33.33);

The INSERT would not work because the R is not in quotes and therefore not a character

d) INSERT INTO ACCOUNT VALUES (2, , , 123.99);

The INSERT would not work because ACCNM cannot be null

*– there **must** be an explanation 1 mark for each*

5 marks

Question A8.

Explain why a Data Warehouse is a good source of data for Data Mining.

DW contains clean, consistent data

DW contains data from multiple sources

DW has inbuilt query capabilities

DW enables DM results to be further investigated at the data source(s)

2 marks per point up to a maximum of 5 marks

5 marks

PART B. Answer any THREE of the following FIVE questions.

Question B9.

- a) A data dictionary or system catalogue is said to contain *meta-data*. Explain the meaning of this term.

Meta-data is data about data i.e. description of the data in the database

(2 marks)

- b) Briefly describe SIX items of information that you would expect to find in a data dictionary

Any six from

Names, types and sizes of data items

Names of relationships

Integrity constraints of the data

Names of authorised users

Access rules for each user on data items

Schema mappings

Usage statistics

Something a bit more than a list ("briefly describe") is expected

(6 marks)

- c) Describe SIX benefits that a data dictionary brings to a database environment.

Any six from

Central control of information about the data resource

Aids user understanding of the data

Enables identification of users who own or can access data

Exact meanings are stored (e.g. is employee equivalent to worker?)

Redundancies/duplications can be identified

Changes to the database can be recorded

The impact of change can be determined

Security can be enforced

Integrity can be ensured

Auditing information can be provided

a brief description is required for the two marks

(12 marks)

TOTAL 20 marks.

Question B10.

a) Briefly define the term data model

A model is an abstraction of complex real-world objects used to build a bridge between the real world and the database

(2 marks)

b) Explain the following terms as they relate to a data model giving a suitable example

i. Entities

Anything about which data are to be collected and stored

ii. Attributes

Characteristics of an entity

iii. Relationships

An association between entities

iv. Constraints

Restrictions placed on data

v. Business rules

An unambiguous description of a procedure or policy in an organisation

1 mark for explanation, 1 mark for a suitable example

(10 marks)

c) Discuss why the Hierarchical and Network models were superseded by the Relational model

Discussion to include

Hierarchical/Network model problems

Requires knowledge of the physical data storage characteristics

They are navigational

Structural change requires a change in all application programs

Other suitable points

Relational DB advantages

Ad hoc queries enabled by SQL

End user isolated from physical-level details

Easier implementation

Easier management

Other suitable points

2 marks per point up to a maximum of 8 marks

(8 marks)

TOTAL 20 marks.

Question B11.

- a) Database performance tuning can be divided into SQL Performance Tuning and DBMS Performance Tuning. Explain the difference between these two aspects of performance tuning.

SQL tuning is client side. Generate a SQL query that returns the correct answer in the least amount of time, using the minimum of resources at the client end typically using indexes, optimisation etc– 4 marks

DBMS tuning is server side. Managing the processes in memory using such techniques as caching and space allocation – 4 marks

(8 marks)

- b) Describe the use of the following in improving performance

- i) Data cache

Storage in primary memory set aside to store data requests so that future requests for data on the same page can be serviced from the cache rather than having to perform another lengthy I/O operation – 4 marks

- ii) SQL cache

Stores the most recently executed SQL statement after it has been through the optimiser so that if the same or a different user issues the same SQL query it will not need to be optimised again – 4 marks

- iii) Sort cache

Used as temporary storage area for ORDER BY or GROUP BY operations. Also used in operations involving an index - 4 marks

(12 marks)

TOTAL 20 marks.

Question B12.

- a) Briefly describe the characteristics of a Distributed Database Management System (DDBMS)

A collection of logically related shared data

Data is split into a number of fragments

Fragments may be replicated

Fragments/replicas are allocated to sites

Sites are linked by a communication network

Data at each site is under the control of a DBMS

DDMS at each site can handle local applications

Each DBMS participates in at least one global application

Other suitable characteristics

2 marks per characteristic up to a maximum of 12 marks

(12 marks)

- b) Explain the difference between homogeneous and heterogeneous DDBMS. (8 marks)

Homogeneous – all sites use the same type of DBMS. Easier to manage, provides incremental growth, increased performance via parallel processing – 4 marks

Heterogeneous – sites may run different DBMS products. Usually brought about by the integration of existing legacy DB. Translations are required to allow communication between different DBMSs – 4 marks

TOTAL 20 marks.

Question B13.

- a) Inmon (1993) defined a data warehouse as a

- i) subject-oriented,

Warehouse is organised around major subjects of the enterprise e.g. customers, products, sales

- ii) integrated,

Source data from different systems are brought together. As this is often inconsistent so different formats must be cleaned up to give the user a consistent view

- iii) time-variant

Data is only accurate at some specific point in time i.e. data stored in the DW represents a snap-shot

- iv) non-volatile

Data is not updated in real-time. New data is added to the historical data in the DW collection of data.

Explain the meaning of each of these terms.

3 marks per term

(12 marks)

- b) Briefly describe FOUR problems that might be associated with developing and managing a data warehouse.

Any four from

Underestimation of time needed to extract, transform and load data into the DW

Problems with the source systems that are only discovered when they are used to feed the DW

Existing OLTP systems do not capture the required data

Increased user demands once the benefits of the DW are noted

Source data that should be regarded as different is categorised as identical when loading onto the DW (data homogenisation)

High demand for resources such as disc space

Conflicts over data ownership. Departments not wanting to store "their" data

High maintenance costs

Long development process, delaying benefits

Difficulty in integrating DW tools to maximise the benefits

Any other suitable point

2 marks per point

(8 marks)

TOTAL 20 marks.

END OF EXAMINATION