IMIS HIGHER DIPLOMA QUALIFICATIONS

DATABASE DESIGN (H4)

Thursday 5th December 2013 10:00hrs – 13:00hrs

DURATION: 3 HOURS

ANSWERS

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.

In the context of the three-level ANSI-SPARC architecture, briefly describe the following

- The external level
- The conceptual level
- The internal level

Explain how this achieves the objective of logical and physical data independence.

(5 marks)

External Level – User view of the data (1 mark)
Conceptual Level – Community view of the data (1 mark)
Internal Level – Physical representation of the data (1 mark)
Explanation of logical independence - changes in conceptual schema doesn't affect external schema (1 mark)
Explanation of physical independence - changes in internal schema doesn't affect conceptual schema (1 mark)

Question A2.

An Object-oriented Data Model (OODM) is the basis for an Object-oriented Database Management System. Define the following OODM terms

- Object
- Class
- Method
- Class Hierarchy
- Inheritance

(5 marks)

Object – an abstraction of a real world entity (1 mark)
Class – a collection of similar objects (1 mark)
Method – an action that can be performed on an object (1 mark)
Class Hierarchy – a tree structure where two or more classes share **one** parent (1 mark)
Inheritance – an object in a class hierarchy inherits the attributes and methods of the classes above it. (1 mark)

Question A3.

Explain the difference between the following SQL data definition commands

- CREATE TABLE
- CREATE TABLE AS
- CREATE VIEW

(5 marks)

CREATE TABLE is standard definition of a table (1 mark)
CREATE TABLE AS will create a new table based on the result of the query that follows AS (2 marks)
CREATE VIEW does something similar but the views are dynamically updated (2 marks)

Question A4.

Explain the concept of referential integrity using a suitable example to illustrate your answer.

(5 marks)

Either the foreign key value in a relation must match a candidate value in its home relation or the foreign key value must be NULL (2 marks)
Suitable example to explain (3 marks)

Question A5.

List FIVE advantages that a Distributed Database Management System (DDBMS) might have over a traditional centralised database system.

(5 marks)

Any five from

- Reflects organisational structure
- Improves shareability
- Improves local autonomy
- Improves availability
- Improves reliability
- Improves performance
- Economies of scale
- Modular growth
- Integration of different systems
- Competitive advantage
- Any other suitable advantage
  (1 mark each to a maximum of five)

Question A6.

Briefly explain FIVE differences between an On-Line Transaction Processing Sy
(OLTP) and a data warehouse.

(5 marks)

Any five from

OLTP	**DW**
Operational	Analytical
Current data	Historic data
Real-time	Not necessarily real-time
Detailed data	Detailed and summarised data
Predictable pattern of updates and queries	Less predictable
High number of transactions	Relatively low numbers of transaction
Predictable and relatively static reporting	Unpredictable, ad hoc reports
Large number of operational users	Smaller number of managerial users

Or any other suitable comparison
One mark per comparison up to a maximum of five.

Question A7.

Define the term cache as used in database performance tuning and explain the difference between the data cache (or buffer cache) and the SQL cache (or procedure cache).

(5 marks)

A cache is a semi-permanent storage area in RAM that holds data from permanent storage – the data files (3 marks)
Data cache – stores the most recently accessed data blocks (1 mark)
SQL cache – stores the most recently executed SQL statements (1 mark)

Question A8.

Explain the difference between a data administrator and a database administrator. Can an organisation make do with just one person for the two roles?

(5 marks)

Data administrator - responsible for the management of the data resource making sure the DB strategy matches the corporate objectives and vision (2 marks)
Database administrator – responsible for the physical realisation of the databases. More technically oriented (2 marks)
May be difficult to find someone with varying abilities but for a small organisation it may be necessary (1 mark)

Total 40 marks

[Turn over]

PART B.
Answer any THREE of the five questions in this section

Question B9.

Dough-Boy is a small bakery company employing 20 people with branches in two locations within a small city. Currently it has separate systems for production control, personnel management, order handling and accounts.

a) By means of suitable examples describe some of the problems that Dough-Boy is likely to face having these disparate systems.

(10 marks)

Answers to include some of the following with suitable examples relating to Dough-Boy

- Separation of data – orders not tightly linked to production
- Duplication of data – order handling and accounts may both hold customer details which can get out of step
- Data dependence – may need to reorganise files for separate systems if the physical structure of one changes
- Incompatible file formats – the systems may not "talk" to each other
- Not good at handling ad hoc queries or reports
- Other suitable points
  (1 mark per point,1 mark per suitable example up to a maximum of 10 marks)

b) The management of Dough-Boy have asked you to prepare a report describing the advantages of using a Database Management System for handling the data requirements of the company. Describe FIVE of these advantages.

(10 marks)

Any five from

- Controlled redundancy
- Consistency of data
- More information from the same data
- Ability to share data
- Improved data integrity
- Enforced standards
- Economies of scale
- Better use of resources, concentration on critical systems
- Improved accessibility of data
- Increased productivity
- Improved maintenance
- Concurrent access possible
- Improved (centralised) back-up and recovery
- Other suitable points

One mark per point and one mark for a suitable description up to a maximum of 10 marks

(Total 20 marks)

Question B10.

Rob, Coronel and Crockett (2008) describe the six phases of the Database Lifecycle, starting with the initial study. Describe the four parts of this initial study, namely

a) Analysis of the company situation

(5 marks)

Look at the company's operating environment (1 mark), what is its mission (1 mark).
Investigate the organisational structure of the company(1 mark). Who reports to whom (1 mark). This helps define information flows, reporting formats etc.(1 mark)

b) Definition of problems and constraints

(5 marks)

Investigate how the existing system functions (1 mark), current inputs (1 mark), current outputs (1 mark), who uses this output (1 mark). Look at informal operations that may vary from the official processes (1 mark)

c) Definition of objectives

(5 marks)

Define the major problems (1 mark), identify common sources of these problems (1 mark), clarify the initial objective of the proposed system (1 mark), consider if it will interface with other systems (1 mark), consider if it will share data with other systems or users (1 mark)

d) Definition of scope and boundaries

(5 marks)

Scope defines the extent of the design (1 mark) and therefore the size of the task e.g. single department or whole of the organisation (1 mark). This helps define the size of the database etc. (1 mark).Boundaries are external limits (1 mark) including technical limitations e.g. using existing hardware (1 mark).

(Total 20 marks)

Question B11.

Dreamboats is an organisation that handles the hiring out of luxery yachts for their owners. Each yacht and its bookings is handled by an agent to ensure the smooth running of the hiring. One of the tables in the database holds details of the yachts that are in the system. This is the CREATE TABLE command used to define the table (N.B. the syntax used is indicative but not meant to represent a specfic database management system. Line numbers have only been added to aid identification)

```
1   CREATE TABLE YachtforHire(
2       YachtNumber       VARCHAR(5)      NOT NULL
3       YachtMake         VARCHAR(15)     NOT NULL
4       CaptainIncluded   CHAR(1)         DEFAULT "Y"
5       NumberofBerths    SMALLINT
6       OwnerNumber       VARCHAR(5)      NOT NULL
7       AgentNumber       VARCHAR(5)
8   PRIMARY KEY (YachtNumber)
9   FOREIGN KEY (OwnerNumber) REFERENCES Owner ON DELETE NO
ACTION ON UPDATE CASCADE
10  FOREIGN KEY (AgentNumber) REFERENCES Agent ON DELETE SET
NULL ON UPDATE CASCADE);
```

a) For line 1 explain the meaning of VACHAR(5)

(2 marks)

A variable number of characters (1 mark) up to a maximum of 5 (1 mark)

b) For line 2 explain the meaning of NOT NULL

(1 mark)

Must contain a valid value (1 mark)

c) For line 4 explain the meaning of CHAR(1)

(1 mark)

A single character can be stored here (1 mark)

d) For line 4 explain the meaning of DEFAULT "Y"

(2 marks)

When a new record is created it will automatically (1 mark) be assigned as the character "Y" but can subsequentkly be overwritten (1 mark)

e) For line 5 explain the meaning of and justify the use of SMALLINT

(2 marks)

Defines the attribute as being able to hold an integer (1 mark). SMALLINT (or even TINYINT) used to reduce the overall record size (1 mark)

f) For line 8 explain the meaning of PRIMARY KEY

(1 ma

Defines the unique identifier to be used for each record (1 mark)

g) For line 9 explain the meaning of FOREIGN KEY…REFERENCES

(2 marks)

Specifies that the value in OwnerNumber for every record in the child table (1 mark) must match with a corresponding record in the parent table, called Owner (1 mark)

h) For line 9 explain the meaning of ON DELETE NO ACTION

(3 marks)

This will prevent deletion of a record from the Owner table (1 mark) if there are matching OwnerNumber values in the YachtforHire table (2 marks)

i) For line 10 explain the meaning of ON DELETE SET NULL

(3 marks)

If a record is deleted from the Agent table (1 mark) the corresponding values in the AgentNumber column of the YachtforHire table will be set to NULL (2 marks)

j) For line 10 explain the meaning of ON UPDATE CASCADE

(3 marks)

If an AgentNumber in the Agent table (1 mark) is updated the corresponding values in the AgentNumber in the YachtforHire table will be updated so that they match (2 marks)

(Total 20 marks)

Question B12.

Transparency is an important aspect of a Distributed Database Management System environment

a) In relation to Distribution Transparency explain the difference between

i. Fragmentation transparency
The highest level of transparency. The end user does not need to know that the DB is partitioned. Fragment names and locations do not need to be specified in the query (4 marks)

ii. Location transparency
The end user knows that the DB is divided into fragments but not where they are held so does not need to specify the location (4 marks)

iii. Local mapping Transparency
The lowest level of transparency. The user needs to specify both the fragment names and the location of the data (4 marks)

(12 marks)

b) In relation to Transaction Transparency describe how a distributed transaction accesses data held at more than one location.

(4 marks)

Each transaction is divided into a number of sub-transactions, one for each site handled by an agent (2 marks). Each sub-transaction can happen simultaneously with the DDBMS ensuring that the whole transaction is completed and transparent to the user (2 marks)

c) Explain in outline how a Distributed Query processor (DQP) works

(4 marks)

DQP decides which fragment to use (1 mark)
If the fragment is replicated it decides which copy/which location (1 mark)
Execution strategy is formed based on costs associated with access time, CPU time and communication costs (2 marks)

(Total 20 marks)

Question B13.

a) ETL is the Extraction Transformation and Loading processes in a data warehouse project. Describe these three processes and some of the issues that need to be considered

(12 marks)

Extraction – targets data sources, typically Online Transaction Processing (OLTP) databases (1 mark) but also spreadsheets, other databases Web logs etc.(1 mark) and extracts this data to temporary storage.(1 mark) Frequency of extractions has to be defined (1 mark)

Transformation – application of a series of rules or functions (1 mark). Data may be merged, split, encoded or summed (2 marks) so that it is clean and consistent with existing data in the DW (1 mark)

Loading – Can be separate from transformation or can be included (1 mark). Constraints applied to ensure integrity such as uniqueness, referential integrity, mandatory fields etc. (2 marks). Frequency of loading has to be determined as it will affect data quality and availability (1 mark)

b) Briefly describe FOUR end-user access tools that could be used in a data warehousing environment

(8 marks)

Any four from

Reporting and query tools – Use of software such as report writers for regular reports, SQL or Query by Example for ad hoc reports (2 marks)

Application development tools – If queries are complex and regular then it might be better to develop an in-house application to resolve the needs of the user of the DW (2 marks)

Online Analytical Processing (OLAP) tools – based on multidimens databases allowing sophisticated analysis of data. Useful for forecasti capacity planning etc. (2 marks)

Data mining tools – used for discovering meaningful patters and trends in the data (2 marks)
Any other reasonable description of an end-user access tool (2 marks)

(Total 20 marks)